Personality, Plasticity and Predictability in Wild White-footed Mice

Natalie Kermany

Thesis submitted to the University of Ottawa
in partial fulfillment of the requirements for the
MSc degree at the
Ottawa-Carleton Institute of Biology

Department of Biology
Faculty of Science
University of Ottawa

Abstract

Recently, there has been an explosion of studies focusing on animal “personality”, defined as consistent individual differences in behaviour. Researchers within this field commonly partition phenotypic variation in behavioural traits at the among- and within-individual levels, because significant among-individual variation implies that a behaviour is repeatable and thus can be subject to selection. It is well known that individuals can not only differ in their mean-level behavioural expression, but also in how their phenotype varies over time, known as temporal plasticity. However, another important yet largely overlooked aspect of behavioural variation is that some individuals are consistently more “unpredictable” than others. This level of variance corresponds to (unexplained) variation among successive measurements on a given individual. Personality and behavioural plasticity have both been studied in many different contexts, however research regarding the possible correlations between these two traits and unpredictability is lacking. Here, I repeatedly assayed locomotor activity in wild white-footed mice (*Peromyscus leucopus*) exposed to a 10-min novel environment test, to determine whether their behavioural reaction norms exhibited consistent among-individual differences in intercept (personality), slope (plasticity), and variance (unpredictability). Using a doubly hierarchical generalized linear mixed model, I found that both the intercept and slope of the temporal reaction norms were significantly repeatable ($R_{\text{int}}= 0.39$, $R_{\text{slp}}= 0.23$) and negatively correlated (r=-0.51). Moreover, unpredictability was found to be significantly repeatable ($R_{\text{sd}}= 0.20$), positively correlated with intercept (r=0.92), and negatively correlated with slope (r=-0.51). To the extent that these correlations reflect the underlying quantitative genetic architecture of behaviour, my results suggest the presence of constraints on the evolution of behaviour across multiple levels of variation.

Resumé

Récemment, il y a eu une explosion d'études portant sur les différences individuelles en comportement, communément appelées la « personnalité » animale. Il est bien connu que les individus peuvent non seulement différer dans leur personnalité (expression moyenne du comportement), mais aussi dans la façon dont leur comportement varie dans le temps, ce qui est connue sous le nom de plasticité temporelle. Cependant, un autre aspect important de la variation comportementale mais largement ignoré est que certains individus sont plus « imprévisibles » que d'autres. Ce niveau de variance correspond à une variation (inexpliquée) entre les mesures successives sur un individu donné. La personnalité et la plasticité comportementale ont toutes deux été étudiées dans de nombreux contextes différents, mais les corrélations possibles entre ces deux traits et l'imprévisibilité restent à faire. J'ai mesuré à plusieurs reprises l'activité locomotrice chez des souris à pattes blanches sauvages (*Peromyscus leucopus*) exposées à un test d'environnement nouveau pour déterminer si leurs normes de réaction comportementale présentaient des différences constantes entre les individus en matière d'intercepte (personnalité), de pente (plasticité) et de variance (imprévisibilité). L'intercepte et la pente des normes de réaction temporelles étaient significativement répétables (R_{int}= 0.39, R_{slp}= 0.23) et négativement corrélées (r=-0.51). De plus, l'imprévisibilité s'est avérée significativement répétable (R_{sd} =0.20), positivement corrélée avec l'intercepte (r=0.92) et négativement corrélée avec la pente (r=-0.51). Dans la mesure où ces corrélations reflètent l'architecture génétique sous-jacente au comportement, mes résultats suggèrent la présence de contraintes sur l'évolution du comportement à travers les niveaux de variation.

Acknowledgements

Firstly, I would like to thank my supervisor Dr. Vincent Careau for giving me the opportunity to work on this project and for accepting me into his lab. From beginning to end, Vince has encouraged me to push boundaries and work hard by leading through example. Thanks for sticking by me throughout the challenging DHGLM coding sessions, and for laughing with me along the way.

I would also like to extend my thanks to the Careau lab, who have helped me at every point along my master’s journey! Their mentorship, empathy, and good humour have always made me feel right at home here in Ottawa. I am so grateful that I joined a lab where I can call my lab mates my good friends. This gratitude also encompasses all of the present and past members who helped collect data in the field which I was gratefully able to use, as well as the Queens University Biological Station for hosting our lab throughout the summers.

I also thank my committee members Dr. Julie Morand-Ferron and Dr. Tom Sherratt for giving constructive feedback and support to this project. A special thank you also goes out to Dr. Julien Martin who provided his valuable expertise and guidance concerning DHGLM modelling and predictability many times over. Additionally, a big thank you to our funding sources (NSERC and the University of Ottawa) whom made this project possible in the first place. I would also like to acknowledge the Anishinaabe, Haudenosaunee, and Algonquin First Nations whose traditional and unseeded territory was essential for the completion of this project.

Lastly, I am forever grateful for the support of my wonderful partner, family, and friends. Even though we are far apart, your support has allowed me to persevere when things were tough. Thank you all for the phone calls in the field, for the constant reassurance, and for the love. With your support I feel like I can accomplish anything the world throws at me!

Contents

1. Introduction

1.1 Animal personality and phenotypic plasticity

Recently, there have been many studies focusing on individual variation in the field of behavioural ecology. Animal "personality", also referred to as "temperament" in the literature, is defined as consistent individual differences in behaviour over time or across environmental contexts (Réale et al., 2007). In this context, "consistent" means that differences in behaviour are maintained when comparing one individual to another, but behaviour is not necessarily fixed within an individual over time. Therefore, the behaviour of an individual can vary during aging or differing environmental conditions, but generally differences between individuals are maintained (Réale et al., 2007).

A common feature of animal personality research is the partitioning of phenotypic variation in behaviour at the among- and within-individual levels (Dingemanse & Dochtermann, 2013). This is because a significant among-individual variation implies that a behaviour is repeatable, and can therefore be potentially heritable and subject to selection (Endler, 1986). This is important because evolutionary biologists are interested in how wild populations evolve over time, and the heritability of a given behaviour is a pre-requisite to any ultimate explanation for the adaptive variation among populations and species. Although behaviour is often regarded as the most labile aspect of an individual's phenotype, most behavioural traits are in fact repeatable (Bell et al., 2009). Yet, a relatively large amount of behavioural variation exists beyond mean individual level expression.

Individuals can not only differ in their overall activity level, but also in how their phenotype varies over time and over environmental gradients – known as phenotypic plasticity

(Montiglio et al., 2010). Personality and plasticity can be studied together through the use of behavioural reaction norms (Araya-Ajoy et al., 2015). While personality specifically is measured through the centered intercept in the behavioural reaction norm, plasticity is measured through the slope representing the extent of change in behaviour. Interestingly, there is increasing evidence to support that personality and plasticity are potentially linked with one another (Dingemanse et al., 2010). For example, less aggressive mice changed their aggression level according to social context (i.e., they were more plastic), however this effect was not observed in more aggressive mice (i.e., they were less plastic) (Natarajan et al., 2009). Behavioural reaction norms can therefore allow the joint analysis of personality and plasticity to effectively study their correlated evolution.

1.2 Predictability

Another important, yet largely overlooked, aspect of behavioural variation is that some individuals are consistently more “unpredictable” than others. This level of variance corresponds to unexplained variation among successive measurements on a given individual (Cleasby et al., 2015; Stamps et al., 2012; Westneat et al., 2015). On a behavioural reaction norm, unpredictability corresponds to the variance in the residuals surrounding an individual’s regression trend line (Stamps et al., 2012). The possibility that individuals consistently differ in predictability is intriguing, as such variation may have an adaptive basis. Heterogeneity in residual variance has been studied in many different morphological traits in the animal breeding literature; however, it is referred to as “dispersion”, “genetic differences in residual variability”, “micro-environmental variance”, “residual variance”, or “micro-environmental sensitivity” (Felleki & Lundeheim, 2015; Hill & Mulder, 2010; Mulder et al., 2009, 2013, 2016; Neves et al., 2012; Rönnegård et al., 2010; Sztepanacz et al., 2017). In these contexts, the heritability for micro-environmental sensitivity can be useful, as it can be selected upon to reduce variance in a desirable trait to optimize profitability, or to predict if a trait's variation will decrease or increase with selection on the mean (Felleki & Lundeheim, 2015). Altogether, studies on morphological traits have shown that there can be a significant heritable component to among-individual variation at the residual level (Hill & Mulder, 2010). Additionally, intraindividual variation (predictability) has been studied in humans, where psychologists have examined traits such as reaction time, emotional stability, reasoning, extraversion, and cognitive skills (Noftle & Fleeson, 2010; Rosengren & Braswell, 2001; Schmiedek et al., 2009; Williams et al., 2005). Interestingly, these studies have found that predictability can change across ontogeny, and that individual differences in predictability can be stable over time and differ between test subjects (Allaire & Marsiske, 2005; Hoffman, 2007; Stamps et al., 2012).

In the context of animal behaviour, however, predictability has rarely been studied, and the heritability of predictability has only been shown in docility (Martin et al., 2017). If predictability in exploratory behaviour were to be repeatable and heritable, it could potentially have implications for predator avoidance, dispersal, or reproductive success (Réale et al., 2007). For example, individuals who are less predictable may have a higher chance of avoiding predators, disperse further in a given environment, or have fewer mating opportunities. For instance, male fallow deer that were either too predictable or unpredictable in their willingness to fight had lower mating success than individuals with intermediate predictability (Jennings et al., 2013). But before we can investigate the adaptive potential of predictability in any behavioural trait, it is crucial to first determine if individual differences in predictability are consistent through time and correlated with personality and plasticity. Although personality and behavioural plasticity are commonly studied together using “random regression” models (Schaeffer, 2004), the addition of a third level of variation (predictability) greatly complicates the statistical procedures involved (Martin et al., 2011). To properly model (co)variance at the mean individual and residual levels, it is necessary to use a doubly hierarchical generalized linear mixed model (DHGLM) (Rönnegård et al., 2010). These DHGLMs are notoriously hard to learn and code, which explains why personality, plasticity, and predictability are rarely studied simultaneously (Westneat et al., 2013). This is unfortunate, because uncovering correlations between personality, plasticity, and predictability could indicate possible constraints on the evolution of behaviour across multiple levels of variation.

1.3 Exploratory behaviour

Personality traits like aggressiveness, sociability, boldness, activity, and exploration are most commonly measured using standardised tests (Réale et al., 2007). Since it was first used on laboratory rats 80 years ago (Hall & Ballachey, 1932), the novel environment test has been heavily used in the fields of psychology and neurobiology to capture aspects of “emotionality”, “exploratory drive”, “anxiety”, or simply locomotion (Archer, 1973; Hall, 1941; Walsh & Cummins, 1976). The novel environment test has recently have gained popularity in behavioral ecology because it was found that the average level of activity expressed throughout a novel environment test can provide insights concerning the behavior of animals in nature (Boon et al., 2008; Boyer et al., 2010; Fraser et al., 2001; Van Overveld & Matthysen, 2010). During exposure to a novel environment, it is well known that activity levels change through time within a single test (Montiglio et al., 2010). These temporal changes in activity are presumably the reflection of short-term habituation to the novel environment – a special case of phenotypic plasticity known as temporal plasticity (Rankin et al., 2009). A behavioural reaction norm approach can capture individual variation in personality and plasticity, with the unexplained variation corresponding to predictability. Therefore, considering minute to minute variation in activity within repeated novel environment tests can allow the joint study of behavioral variation across distinct levels (personality, plasticity, and predictability).

Though the behavioural reaction norm approach is useful to study the combined variation in personality, plasticity, and predictability, there are implications for the variable selected to quantify exploration. Studies using open-field tests have used a variety of variables to quantify exploratory behaviour, such as distance moved in the open-field, number of gridlines crossed, proportion of the arena covered, or the duration of time spent in the centre vs periphery (Careau et al., 2012; Elder et al., 2008; Houslay et al., 2018; Montiglio et al., 2010; Szechtman et al., 1994; Walsh & Cummins, 1976; Warren & Callaghan, 1975). In order to construct the behavioural reaction norms, it is necessary to use a continuous variable for the y-axis, which complicates the use of proportion variables, and leaves distance moved as a viable option that can be readily generated. Focusing on distance moved exclusively to quantify exploration can be criticized, as individuals with a high activity level in the open-field may be moving in only one area of the arena (thus not exploring the majority of space), and yet are still considered more ‘explorative’ in this context. It is therefore important for studies focusing on exploration to consider the relevance of their variable used to quantify the behaviour, in order to make reasonable inferences on the personality trait. In this study, distance moved in the open-field was used to quantify locomotor activity in a novel environment, which best allows for the combined study of personality, plasticity, and predictability in the context of a novel environment.

1.4 Objectives

Personality and plasticity have both been studied in many different behavioural contexts and in numerous species; however, research regarding the possible correlations between these two traits and unpredictability is scarce. Additionally, among-individual differences in unpredictability have yet to be shown in exploratory behaviour. This project will attempt to fill these gaps through modelling repeated behavioural reaction norms to estimate the repeatability of personality (intercept), plasticity (slope), and unpredictability (variance). Through modelling all three levels of variation simultaneously, I will also attempt to find whether unpredictability in exploratory behaviour is correlated with the mean individual-level expression and plasticity of exploratory behaviour.

2. Materials and Methods

2.1 Study species

This project used wild white-footed mice (*Peromyscus leucopus*) as a study model. White-footed mice are beneficial to use as they are both easy to catch, as well as abundant at the study site, allowing for adequate sample size. *Peromyscus leucopus* are nocturnal mice that have a polygynandrous mating system, and a home range that covers ~580m^2 according to population density estimates (Wolff, 1985, 1996; Xia et al., 1991). Their populations fluctuate with annual mass production, and they are more abundant during high food availability periods (mass years) (Wolff, 1996). Additionally, male home ranges are two times larger than female home ranges and overlap, while female home ranges are non-overlapping as breeding individuals become territorial to immigrating females (Vessey & Vessey, 2007; Wolff, 1996).

2.2 Study site

Field work for this project was conducted at the Queens University Biological Station located 50km north of Kingston, Ontario (44°33’5”N; 76°19’27”W), and data collection for this project was authorized by the Ontario Ministry of Natural Resources. Two trapping locations were used on site, termed Cow Island and Blueberry Hill. Both trapping sites were covered by deciduous forest, populated primarily with sugar maple (*Acer saccharum*) and oak trees (*Quercus rubra and Q. alba*). Each trapping site contained a 15 × 15 m sampling grid that alternated between Longworth-style traps (box: 8.57 x 6.35 x 13.97 cm; tunnel: 4.45 x 4.76 x 9 12.7 cm) and nest boxes. However, only Longworth-style traps were used for capturing individuals in this study. In total, 181 Longworth traps were used to trap individuals daily during the summer months over the course of four years (2016-2019).

2.3 Captures

All mice were captured and handled in accordance to protocols approved by the University of Ottawa Animal Care Committee and the Queen’s University Animal Care Committee. Longworth traps were set at dusk in the evening and were baited with sunflower seeds, a small piece of apple, and a square piece of polyester batting for insulation. Each trap was checked the following morning at dawn. In cases where an individual was captured for the first time, they were assigned a unique identification code and a permanent identification tag was installed on each ear. On each capture, the body mass, sex, age, reproductive status, and parasite count of the mouse were recorded. Age was recorded as juvenile or adult according to colour patterns observed on the pelage. Juvenile mice exhibited brown and grey pelage above the molt line, while adults were completely brown above the middorsal molt line (Collins, 1923). Reproductive status was categorized as active or not depending on external sexual characteristics. After these steps were completed at the study site, individuals were either released or placed back in their trap and transported to a laboratory located ~600m away for behavioural measurement. Up to 16 mice were transported to the laboratory each day, and mice that had been tested less than 7 days prior were released at the capture site. Moreover, females that were categorized as pregnant or lactating were also released at the capture site and did not undergo behavioural testing. All mice were released at their capture location on the same day.

2.4 Behavioural measurement

Behavioural measurement in the laboratory consisted of an open-field test, which involved placing a mouse into a circular arena (radius = 70 cm) surrounded by black curtains, and video recording its movement for 10 minutes (see Figure 1). The arena used in the test simulates a relatively large novel environment where escape is prevented by surrounding walls (86 cm high). Each test was recorded using a camera (Basler ace IR) positioned 2.09m above the arena and connected to the Ethovision^{XT} software. Tracking immediately began when the mouse was detected inside the arena, and continued for 600 seconds at which point the test ended. The mouse was then removed from the arena and weighed using a precision scale (Mettler Toledo, Model ML1602T/00). The arena was washed with Accel® disinfectant and rinsed with water before and after each test. The Ethovision^{XT} software was used to extract the distance moved by the mouse for each minute of the test, yielding 10 values for each individual. These values were then used to estimate the behavioural reaction norms of distance moved in the open-field over time.

2.5 Statistical analyses

I used a DHGLM to simultaneously analyze behavioural (co)variation at the mean (personality and plasticity) and residual levels (unpredictability) in a single step. The DHGLM was run using the brms package which fits Bayesian multilevel models in R (version 3.6.2) using the programming language Stan (Buerkner, 2017), and models the standard deviation of the behavioural reaction norms. Locomotor activity (distance moved in the open-field), the response variable, was square root transformed to correct for overdispersion, and scaled (mean = 0, variance = 1) prior to analysis. The fixed effects included in the mean and dispersion parts of the model were test minute, Julian day, time of day, test sequence, sex, age, body mass, and reproductive status. Test minute squared (after centering test minute) was also included in the model to account for the non-linearity of the reaction norms (Figure 2). Therefore, each variable (e.g., sex) appeared twice in the model: in the mean part of the model to test for an influence on the intercept of the reaction norm (personality) and again in the dispersion part of the model to test for an effect on residual variance (unpredictability). To test if a given variable also influenced the slope of the reaction norms (plasticity), interactions between test minute and each main effect above in the mean part of the model were included. All continuous variables were z-transformed (mean = 0, variance = 1) and all categorical variables (sex, age, reproductive status) were centered to allow for proper interpretation of fixed effects when involved in interactions (Schielzeth, 2010).

Following Araya-Ajoy et al. (2018), a multilevel approach was adopted to analyze the repeated behavioural reaction norms and partition the phenotypic variation in intercept (personality) and slope (plasticity) at the among- and within- individual levels. Moreover, the multilevel approach was extended to the dispersion part of the model, allowing to effectively partition variance in the residuals (unpredictability) at the among- and within-individual levels. Two random effects were included in the model, both fitted to include a random intercept term and a random slope term to the mean part of the model, and random intercept to the dispersion part of the model, and their covariance (i.e., 3x3 matrix). The first random effect included was individual identity ("ID") to account for the among-individual variation in intercept, slope, and variance. The second random effect was a variable grouping all 1-min observations made on a given individual within a given test ("series") to account for the within-individual variation in intercept, slope, and variance. The repeatability of personality, plasticity, and unpredictability were calculated using the following equations respectively:

(a) $V_{\text{int among}} / (V_{\text{int among}} + V_{\text{int within}})$

(b) $V_{\text{slope among}} / (V_{\text{slope among}} + V_{\text{slope within}})$

(c) $V_{\text{sd among}} / (V_{\text{sd among}} + V_{\text{sd within}})$

The DHGLM was fitted using 3 Markov chains, each including 50 000 iterations, 10 000 burn in iterations, and a thinning interval of 40. Default priors were used on all parameters, which specified a half-*t* student distribution on the random effects, with 3 degrees of freedom and a standard deviation of 10. Moreover, an LKJ correlation matrix with 1 degree of freedom was specified, and acted as a flat prior on the correlations. The DHGLM converged, as indicated by Rhat values (range 0.9994-1.0026), effective sample sizes measures (Bulk >2641, Tail >2432), and assessment through visually inspecting the trace plots. Moreover, autocorrelation among successive samples was checked to ensure that all parameter values fell below 0.05, and a maximum value of 0.02 was indicated.

3. Results

3.1 Descriptive statistics

A total of 909 open-field tests were conducted on 436 individuals (200 females and 236 males) captured over the course of four field seasons (2016-2019). On average, individuals were brought back to the lab twice, ranging from one to ten visits throughout the season. Only two individuals were tested in two different years.

3.2 Population-level parameters

3.2.1 Reaction norm intercept (average activity)

Adults on average exhibited a lower intercept on their behavioural reaction norms compared to juvenile mice (Table 1). Mice who were tested later in the season were less active than individuals tested earlier (Table 1). Test sequence also had a significant and negative effect on locomotor activity, reflecting that individuals decreased in their distance moved in the open-field with each consecutive test (Table 1). Lastly, mice tested later in the day were less active compared to those tested earlier (Table 1). Sex, body mass, and reproductive status did not exert an influence on the intercept of the behavioural reaction norms (Table 1).

3.2.2 Reaction norm slope (plasticity)

As expected, test minute had a significant and negative effect on locomotor activity during the open-field test, such that individuals on average decreased in their distance moved over time (Table 1; Figure 2). Interactions with test minute were included in the model to identify parameters that had an effect on the slope of the behavioural reaction norms. The interaction with Julian day was significant and had a positive estimate, such that individuals tested later in the season had shallower (less negative) slopes (Table 1). Similarly, the interaction with test sequence was significant and had a positive estimate, such that the slopes became shallower as individuals were repeatedly tested (Table 1). Lastly, individuals that were heavier on average had slopes that were steeper (more negative) than lighter individuals (Table 1). Time of day, sex, age, and reproductive status did not influence the slope of the behavioural reaction norms (Table 1).

3.2.3 Residual variance (unpredictability)

Test minute had a significant and negative effect on the variance of the reaction norms, meaning that, on average, individuals became more predictable as the open-field test progressed (Table 1). Similarly, mice that were tested later in the season were on average more predictable than mice tested earlier (Julian day; Table 1). Lastly, predictability increased on average with each consecutive visit to the lab (test sequence; Table 1). Time of day, sex, age, body mass, and reproductive status had no significant effect on residual variance (Table 1).

3.3 Among- and within-individual (co)variance

Both the intercept and slope of the behavioural reaction norms were significantly repeatable with estimates of R = 0.39 and 0.23, respectively, and negatively correlated at the among-individual level (Table 2). Therefore, individuals with higher activity on average throughout the open-field test showed a greater decrease in their activity over time (Figure 3A). Unpredictability (residual standard deviation) was found to be significantly repeatable (Table 2) with an estimate of R = 0.20. At the among-individual level, unpredictability was positively correlated with the intercept ($r_{sd.int}$=0.92[0.64,1.00]; Figure 3C), implying that that individuals with a higher activity on average had a higher variance exhibited on their behavioural reaction norms. Moreover, unpredictability was negatively correlated with slope ($r_{sd.slp}$=-0.51[-0.79, -0.12]; Figure 3E), meaning that individuals with steeper (more negative) slopes were also less predictable.

At the within-individual level, the intercept and the slope of the behavioural reaction norms were negatively correlated (r_{series}= -0.19[-0.32, -0.07]; Figure 3B). Moreover, unpredictability was positively correlated with intercept (r_{series}= 0.26[0.13,0.34]; Figure 3D). Therefore, when an individual displayed a higher activity than usual, its activity also decreased more than usual (slope) and it was less predictable than usual. There was no significant relationship between slope and unpredictability at the within-individual level (Table 2; Figure 3F).

4. Discussion

4.1 Overview

The objectives of this study were to 1) model repeated behavioural reaction norms to estimate the repeatability of personality, plasticity, and predictability and 2) uncover the potential correlations across these three levels of variation at the among- and within- individual levels. Overall, I found that individuals significantly differed in their intercept, slope, and variance of their behavioural reaction norms. Furthermore, unpredictability was negatively correlated with intercept, and positively correlated with slope at the among-individual level. This study demonstrates that in the context of a novel environment test, individuals can not only differ in their mean activity level expression, but also in the variance of their activity expression. To the extent that the among-individual correlations reflect the underlying quantitative genetic architecture of behaviour, my results suggest constraints on the evolution of behaviour across multiple levels of variation. To my knowledge, this study is also one of the first to determine the population-level parameters that influence predictability of exploration in a wild population.

4.2 Covariates of personality, plasticity, and predictability

Personality

Average distance moved in the open-field was influenced by age, Julian date, test sequence, and time of day. Adults on average moved less in the open-field compared to juvenile mice, which is consistent with findings in other species of mice (*Micromys minutus, Mastomys natalensis*) (Schuster et al., 2017; Vanden Broecke et al., 2018), and a variety of other taxa (Biondi et al., 2013; Miller et al., 2015; Ray & Hansen, 2005). Differences in exploration between adults and juveniles has been attributed to the need for juveniles to gather more information about their environment early in life, as well as their lack of experience with predators compared to adults (Reader, 2015; Rödel et al., 2015). Moreover, it has also been suggested that juveniles are less efficient at gathering information compared to adults, and therefore require more time exploring novel environments to gather the equivalent amount of information (Biondi et al., 2013). The negative effect of Julian date on distance moved means that individuals tested later in the season were less active than individuals tested earlier (controlling for test sequence; see below), and this seasonal pattern might be related to ageing effects within the coarse age categories (i.e., juvenile/adult).

The negative effect of test sequence reflects that individuals decreased in their distance moved with each consecutive visit to the lab, which is a phenomenon that has been found in many exploration studies (Archer, 1973; Dingemanse et al., 2002; Martin & Réale, 2008). A negative test sequence effect suggests a long-term habituation effect over the season, where individuals effectively gathered knowledge based on previous testing experience, leading to decreased activity in the open-field in subsequent testing.

Lastly, time of day had a significant and negative effect on average locomotor activity, indicating that individuals who were tested later in the day were less active on average compared to those tested earlier. This was somewhat expected, as white-footed mice are nocturnal mammals who are most active at night and the open-field testing occurred primarily in the morning. However, one confounding factor is the time spent in captivity before the open-field testing. Mice that were tested later in the day spent more time in the trap between time of capture and testing, and this could have contributed to those individuals moving less in the open-field compared to individuals who spent less time in captivity.

Plasticity

Interactions with test minute showed which factors influenced the slope of the behavioural reaction norms, and consequently short-term habituation to the open-field. Individuals who were tested later in the season had shallower slopes compared to those tested earlier, and individuals also displayed shallower slopes upon repeated testing. These results are consistent to those found in eastern chipmunks, where individuals who were tested a second time in the open-field became less active, with shallower slopes (Montiglio et al., 2010). A negative test sequence effect on plasticity is also suggesting of a long-term habituation effect to the open-field. Rankin et al. (2009) defines the first characteristic of habituation as repeated application of a stimulus resulting in a progressive decrease of a response, where the change may include decreases in the magnitude of a response. This is reflected in my study, as the repeated stimulus is the repeated open-field testing, and the decrease in magnitude of a response is the decrease in overall activity (i.e., intercept; see above) and plasticity (i.e., shallower slopes). Interestingly, mice that were heavier on average had slopes that were steeper (more negative) than lighter individuals during open-field testing. This result is opposite to a previous finding in humans, where individuals with a lower body mass index habituated faster (i.e., steeper slope) to psychological stress, compared to those who with a high body mass index (Feda et al., 2015). Aside from this study on humans by Feda et al., I am not aware of another study finding an effect of mass on habituation.

Predictability

The predictability of individuals during open-field testing was influenced by test minute, Julian date, and test sequence. Both test minute and test sequence had significant and negative estimates, which is indicative of both short- and long-term habituation to the open-field. Negative test minute effects indicate short-term habituation within a test: mice became more predictable as they gathered more information through the progression of the test. By contrast, test sequence effects indicate long-term habituation across the season: mice became more predictable with each consecutive open-field test. A negative test sequence effect is consistent with empirical evidence from the psychology literature, where the variance of individuals declines with increased experience and development (Stamps & Krishnan, 2014). In these theoretical models, learning follows a Bayesian updating process where the past experience of an individual informs their future decisions, and this causes individuals to have a decreased need to be highly variable when the same situation is encountered (Stamps & Krishnan, 2014). Julian date also had a negative effect on the variance, meaning that individuals who were tested later in the season were more predictable compared to individuals who were tested earlier. This seasonal effect again might be related to ageing effects within the coarse age categories (i.e., juvenile/adult). There could also be potential seasonal/environmental factors that were not taken into account in the model that could have led to individuals being more predictable later in the season compared to earlier.

4.3 Repeatability

Accounting for the fixed effects mentioned above, the repeatability of intercept was estimated as $R = 0.39$, meaning that 39% of the total phenotypic variation in intercept is attributed to differences among individuals. This repeatability estimate is consistent with the literature, as a meta-analysis over a range of taxa found the average repeatability estimate to be $R = 0.37$ in behavioural studies (SE = 0.01, N studies = 759, N taxa = 98) (Bell et al., 2009). More specifically, studies focusing on exploration in species of mice and chipmunks have reported repeatabilities ranging from $R = 0.30$-0.54, which are comparable to my study (Careau et al., 2015; Montiglio et al., 2010; Newar & Careau, 2018; Schuster et al., 2017; Vanden Broecke et al., 2018). It is important to note however, that most repeatability estimates in the behavioural literature are calculated for the actual behavioural traits and not reaction norm components (i.e., intercept), and therefore include residual error in their calculation (V_e) (Araya-Ajoy et al., 2015; Sonnweber et al., 2018). While trait repeatability is defined as the proportion of phenotypic variance that is attributed to among-individual differences, reaction norm repeatability represents how stable individuals are in their responses to environmental change over time (Sonnweber et al., 2018). When more accurately comparing my repeatability results with another behavioural reaction norm study on aggressiveness, I found that my estimate of intercept repeatability was considerably lower than their intercept estimate of $R=0.57$ (Araya-Ajoy & Dingemanse, 2017).

In terms of the repeatability estimate of slope, 23% of the variance was attributed to among-individual differences. A different study on aggressiveness in male great tits reported a higher repeatability of slope compared to my study (R=0.52), and used the same multilevel reaction norm method (Araya-Ajoy & Dingemanse, 2017). There is substantial variation in the duration of open-field tests, which can in turn impact estimates of repeatability and lead to disparities among studies (Montiglio et al., 2010). Studies on rodents specifically have used various cut-offs ranging from 4.5 to 60 minutes (Boon et al., 2007; Careau et al., 2015; Elder et al., 2008; Montiglio et al., 2010; Newar & Careau, 2018; Schuster et al., 2017; Szechtman et al., 1994; Vanden Broecke et al., 2018). In my study, the cut-off time for the open-field test was 10 minutes; had we use a 5 min cut-off, we would have obtained a repeatability of 0.48 [0.30,0.65] for the slope (see table S1 in Appendix), thus illustrating the important point raised by Montiglio et al. (2010).

Another aspect that varies greatly in the animal personality literature is the way in which studies have calculated repeatabilities. In my approach, I applied the same formula (see equations (a), (b), and (c) in section 2.5 above) to calculate repeatability for the intercept, slope, and variance of the reaction norms. In order to use this approach, it is necessary to apply statistical methods (i.e., mixed models) that partition the overall phenotypic variation to the among- and within-individual levels. Some studies have used this multilevel approach in calculating the repeatability of slope (Araya-Ajoy et al., 2015; Araya-Ajoy & Dingemanse, 2017; Careau et al., 2020; Montiglio et al., 2010). The advantages of a multilevel approach to calculating repeatability is that it provides both a reliable and robust estimate that is comparable across the literature in other behavioural studies (Araya-Ajoy et al., 2015). Despite these major advantages however, no other study to date has used this approach in calculating the repeatability of predictability. One major challenge in adopting this approach is the amount of required data and repeated measures, which is often too large for many researchers and is not feasible under time constraints (Mitchell et al., 2016). In my study however, the use of video-tracking software to break up the average distance moved for each minute of the test allowed me to have the necessary amount of repeated measures (i.e, 10 per test) to utilize this multilevel framework. Using this approach, I found that unpredictability was significantly repeatable, with 20% of the variance attributing to among-individual differences. This result is exciting, as it provides the first evidence of predictability being repeatable in exploratory behaviour using these methods. Repeatability is the first criteria for selection to act on a trait, and I suggest that future research should investigate whether predictability in exploration meets the other criteria for selection: heritability, and effects on fitness.

4.4 Hierarchical correlations

The second objective of this study was to uncover possible correlations among personality, plasticity, and predictability. As expected, the intercept and slope of the behavioural reaction norms were negatively correlated, such that individuals who were more active on average exhibited steeper, more negative slopes. This study is not the first to find a relationship between personality and plasticity (Araya-Ajoy & Dingemanse, 2017; Biro et al., 2013; Montiglio et al., 2010; Natarajan et al., 2009; Nussey et al., 2005; Westneat et al., 2011). A negative covariance between the intercept and the slope during open-field testing has been described as evidence for two distinct methods for exploration: 1) superficial and 2) thorough (Montiglio et al., 2010). The first method is used by individuals who explore their environment 'superficially' through being more active at the beginning of the test, then rapidly decreasing in their activity over time (i.e., high intercept and steep slope) (Montiglio et al., 2010). The second group of individuals are grouped as 'thorough' explorers exhibiting consistent movement throughout the test with little change (i.e., low intercept and shallow slope) (Montiglio et al., 2010).

A negative intercept-slope covariance may also arise through a common statistical phenomenon known as 'regression towards the mean'. Regression towards the mean occurs when two related measurements are taken from a distribution, and when the first measurement is extreme, the second measure falls closer to the population mean by chance (Nesselroade et al., 1980; Stigler, 1997). This means that when repeated measures are taken on the same individual, natural variation can appear as a significant negative correlation between an individual's state at time 1 and the change in state from time 1 to time 2 (Barnett et al., 2005; Kelly & Price, 2005). Considering that regression towards the mean can occur in any variable that has a randomly distributed error, it is important to rule this out when evaluating if the intercept-slope covariance is a biological feature of the behavioral reaction norms (Barnett et al., 2005). One of the statistical ways to control for regression towards the mean is the use of ANCOVA analysis, as each subject's follow up measurement is adjusted according to their baseline measurement (Barnett et al., 2005). Analysis of covariance is effectively done through the use of the DHLGM, as each individual tested is categorized into a unique cluster, and then has their second test minute measure adjusted according to their cluster's test average (calculated from the total 10 measurements within that given test). This means that for every individual tested in the open-field, their subsequent measures of activity are corrected by their individual test average, which successfully takes into account the differences caused by regression towards the mean.

One of the other major strengths of the DHGLM analysis is the ability to analyze personality, plasticity, and predictability simultaneously to determine their covariance, which is preferable to two-step approaches involving extracting the residuals from a random regression model and calculating variance in the residuals for each individual (Biro & Adriaenssens, 2013; Briffa, 2013; Chang et al., 2017; Jennings et al., 2013; Jolles et al., 2019; Stamps et al., 2012). Through the use of the DHGLM, unpredictability was found to be positively correlated with intercept, and negatively correlated with slope at the among-individual level. This means that individuals who had a high activity on average were less predictable compared to those with low activity. Moreover, individuals who were less plastic were more predictable than those who were more plastic.

Finding a positive relationship between the intercept and variance of the reaction norms is somewhat inconsistent with the literature available on predictability. Mitchell et al. (2016) found a negative correlation between mean activity rate and unpredictability in hermit crabs, which is opposite to the relationship I found. In a different study on hermit crabs, Stamps et al. (2012) found a significant positive relationship between average latency to emerge and unpredictability, again contrasting to my results. Lastly, a study on mosquitofish found that there was no relationship between unpredictability and average activity (Biro & Adriaenssens, 2013). There is however, one study in agreement with my results. In agama lizards, Highcock and Carter (2014) found that flight initiation distance was negatively correlated with unpredictability, implying that bold individuals were less predictable than shy individuals. This is in agreement with my results, as individuals who were the most 'explorative' in the open-field were also the least predictable. One possible reasoning for this is that individuals can offset the risk of being more explorative by being less predictable in their exploratory behaviour (Highcock & Carter, 2014). This is based on the assumption that protean escape behaviour (unpredictable movement) is predicted to decrease predation risk, as predators experience difficulty anticipating the future position of their prey target (Driver & Humphries, 1970; Jones et al., 2011; Richardson et al., 2018). Moreover, this logic also agrees with the previous claim that unpredictability increases with predation risk (Briffa, 2013). In order to investigate these possibilities further, I suggest future research could investigate how predictability of exploration varies in open-field tests with the presence and absence of predator cues (e.g., scent, sounds, displays).

In terms of the relationship between unpredictability and slope, they were negatively correlated such that individuals who were more plastic on average were less predictable. This is in line with the predictions that individuals with increased plasticity should also exhibit increased variance due to higher organismal error in calculating their ideal trait values (DeWitt et al., 1998; Mitchell et al., 2016; Tonsor et al., 2013; Westneat et al., 2015). Another explanation that has been proposed is that responsive (i.e., more plastic) individuals should select for unpredictable behavioural patterns when predictable individuals can be exploited (Wolf et al., 2011). Some of the situations in which predictable individuals can be exploited are prey species avoiding exploitation from predators, or foraging species that avoid exploitation of their search effort and information from competitors (Wolf et al., 2011). In my case, individuals being more unpredictable in the way they explore may serve as a decreased risk to predation, potentially driving this slope-variance correlation. It will therefore be important to consider the relevance of exploration behaviour in relation to exploitation, in order to interpret a negative plasticity-predictability relationship. Lastly, this study also demonstrates that separating co-variance at the among- and within-individual levels is important, considering that plasticity was only correlated with predictability at the among- but not within-individual level.

Together, these correlations between personality and plasticity to predictability suggest that individuals are constrained not only to an ‘optimal’ reaction norm shape, but also a specific variance. Furthermore, through expanding the DHGLM framework to quantify these correlations, we can better evaluate whether personality, plasticity, and predictability are potentially involved in a combined behavioural syndrome (Mitchell et al., 2016). However, this study on individual variation is only the first step towards a better understanding of the quantitative genetic architecture of behaviour across multiple levels of variation. To uncover the true genetic (co)variance, future studies require a population with a known pedigree, and can therefore predict how selection on one level (i.e., personality) impacts the other levels in question (i.e., plasticity and predictability).

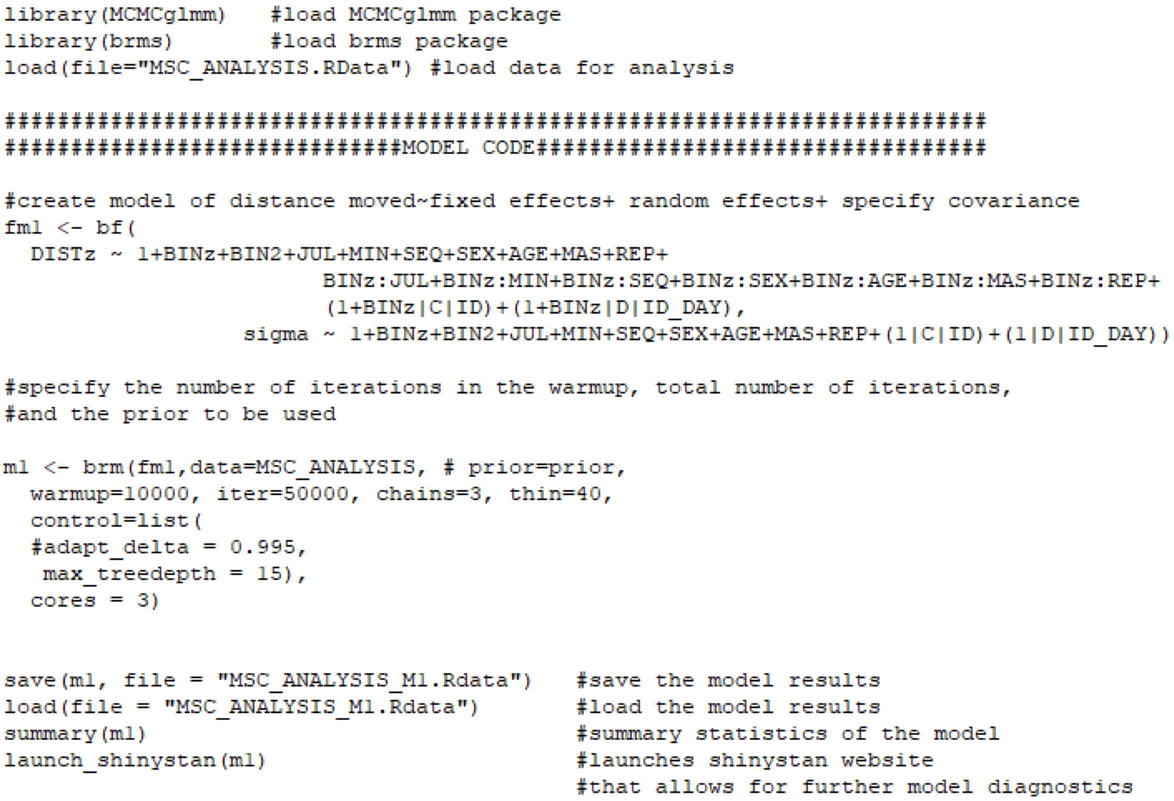
Another aspect of this study that can be improved on in future research is the model specification. In the current DHGLM analysis, the response variable (distance moved in the open-field) was transformed (square root) and standardized to have a mean = 0 and variance = 1. However, a more accurate model would specify the response variable to have a truncated Gaussian distribution with a lower limit of 0 (as individuals could not move less than 0cm in the open-field). This additional truncation specification would ensure the most accurate correlations from the model and avoid any potentially spurious results from the response variable being bound to 0 (Lee et al., 2003; Smith et al., 2009). When attempting to fit this truncated DHGLM to my dataset, it failed to converge due to the inadequate sample size as well as the existing complex model structure. It is therefore important for future studies to consider the amount of necessary data required to run these complex DHGLM models, as well as understand how the distribution of the response variable can limit the available methods for analysis. Though the truncated DHGLM is the gold-standard for the most accurate correlation results, the current DHGLM analysis is a significant improvement from previous two-step approaches used in the existing literature, and therefore serves as an important step forward in understanding and analysing behavioural predictability.

5. Conclusion

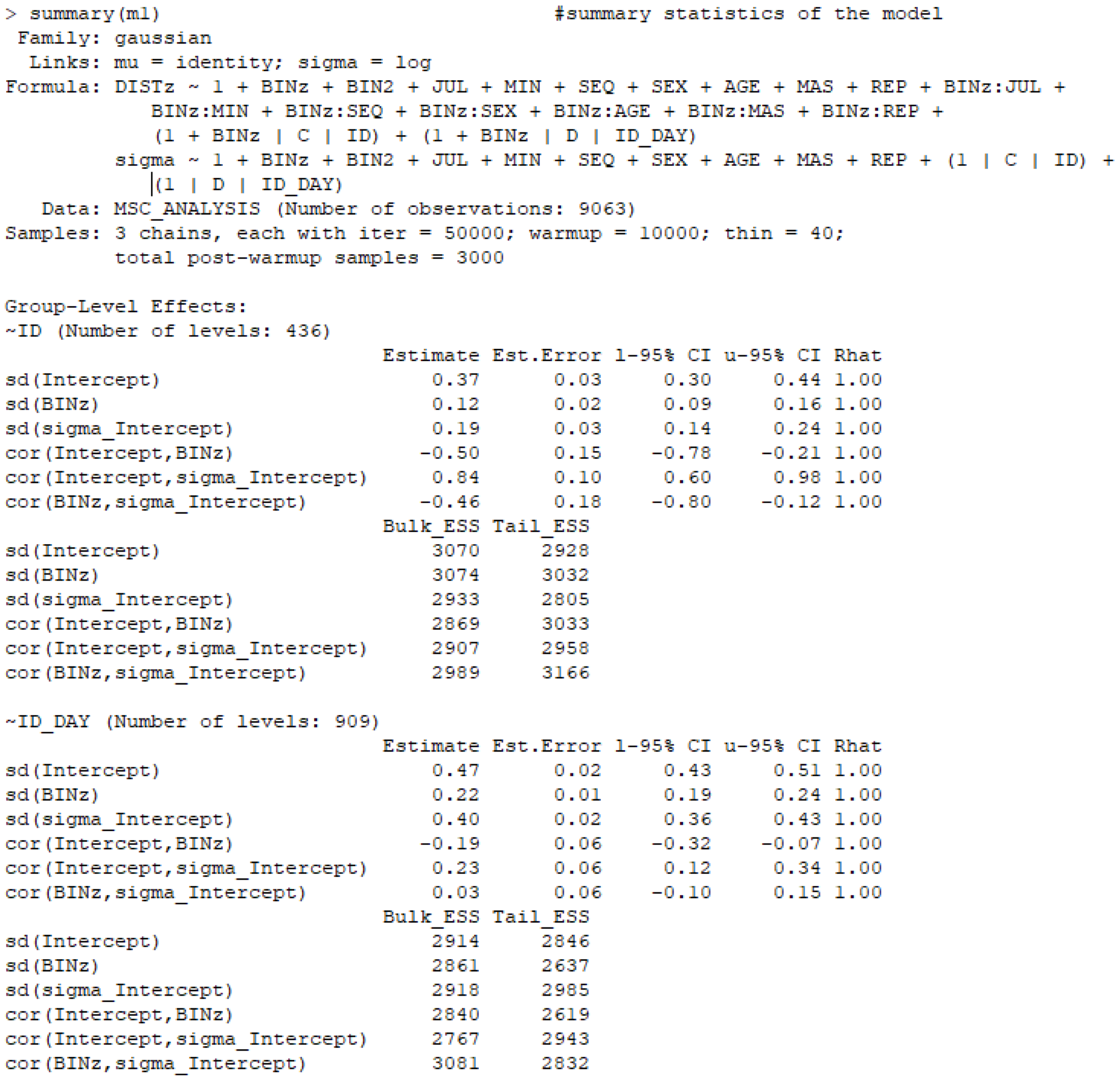
It has been well established that individuals can differ both in their mean phenotype (personality), and how their phenotype varies over time (plasticity). However, research on personality and plasticity has rarely considered the unexplained variation around behavioural reaction norms. This study contributes to fill in this research gap by determining if individuals repeatedly differ in their ‘unexplained’ behavioural variation in the context of a novel environment test. I found that the intercept, slope, and variance of the behavioural reaction norms were significantly repeatable. Moreover, unpredictability was positively correlated with personality, and negatively correlated with plasticity. This study demonstrates that through a multilevel DHGLM framework, the calculation of the repeatability of predictability can be done, and that these three levels of behavioural variation are constrained to a certain extent. I suggest that future research focus on investigating if predictability 1) is heritable as it is in docility (Martin et al., 2017), 2) has fitness consequences, and 3) correlates with other relevant aspects of the phenotype, such as locomotor performance (Berberi & Careau, 2019) and metabolic rate (Fiedler, 2019).

6. R Markdown

This section shows the R code used to run the DHGLM, as well as the model output.



```
library(MCMCglmm)   #load MCMCglmm package
library(brms)       #load brms package
load(file="MSC_ANALYSIS.RData") #load data for analysis

##########################################################################
##############################MODEL CODE##################################

#create model of distance moved~fixed effects+ random effects+ specify covariance
fml <- bf(
  DISTz ~ 1+BINz+BIN2+JUL+MIN+SEQ+SEX+AGE+MAS+REP+
                        BINz:JUL+BINz:MIN+BINz:SEQ+BINz:SEX+BINz:AGE+BINz:MAS+BINz:REP+
                        (1+BINz|C|ID)+(1+BINz|D|ID_DAY),
                  sigma ~ 1+BINz+BIN2+JUL+MIN+SEQ+SEX+AGE+MAS+REP+(1|C|ID)+(1|D|ID_DAY))

#specify the number of iterations in the warmup, total number of iterations,
#and the prior to be used

ml <- brm(fml,data=MSC_ANALYSIS, # prior=prior,
  warmup=10000, iter=50000, chains=3, thin=40,
  control=list(
  #adapt_delta = 0.995,
   max_treedepth = 15),
  cores = 3)


save(ml, file = "MSC_ANALYSIS_M1.Rdata")   #save the model results
load(file = "MSC_ANALYSIS_M1.Rdata")       #load the model results
summary(ml)                                #summary statistics of the model
launch_shinystan(ml)                       #launches shinystan website
                                           #that allows for further model diagnostics
```



```
> summary(m1)                                #summary statistics of the model
 Family: gaussian 
  Links: mu = identity; sigma = log 
Formula: DISTz ~ 1 + BINz + BIN2 + JUL + MIN + SEQ + SEX + AGE + MAS + REP + BINz:JUL + 
            BINz:MIN + BINz:SEQ + BINz:SEX + BINz:AGE + BINz:MAS + BINz:REP + 
            (1 + BINz | C | ID) + (1 + BINz | D | ID_DAY) 
         sigma ~ 1 + BINz + BIN2 + JUL + MIN + SEQ + SEX + AGE + MAS + REP + (1 | C | ID) + 
            (1 | D | ID_DAY)
   Data: MSC_ANALYSIS (Number of observations: 9063) 
Samples: 3 chains, each with iter = 50000; warmup = 10000; thin = 40;
         total post-warmup samples = 3000

Group-Level Effects: 
~ID (Number of levels: 436) 
                               Estimate Est.Error l-95% CI u-95% CI Rhat
sd(Intercept)                      0.37      0.03     0.30     0.44 1.00
sd(BINz)                           0.12      0.02     0.09     0.16 1.00
sd(sigma_Intercept)                0.19      0.03     0.14     0.24 1.00
cor(Intercept,BINz)               -0.50      0.15    -0.78    -0.21 1.00
cor(Intercept,sigma_Intercept)     0.84      0.10     0.60     0.98 1.00
cor(BINz,sigma_Intercept)         -0.46      0.18    -0.80    -0.12 1.00
                               Bulk_ESS Tail_ESS
sd(Intercept)                      3070     2928
sd(BINz)                           3074     3032
sd(sigma_Intercept)                2933     2805
cor(Intercept,BINz)                2869     3033
cor(Intercept,sigma_Intercept)     2907     2958
cor(BINz,sigma_Intercept)          2989     3166

~ID_DAY (Number of levels: 909) 
                               Estimate Est.Error l-95% CI u-95% CI Rhat
sd(Intercept)                      0.47      0.02     0.43     0.51 1.00
sd(BINz)                           0.22      0.01     0.19     0.24 1.00
sd(sigma_Intercept)                0.40      0.02     0.36     0.43 1.00
cor(Intercept,BINz)               -0.19      0.06    -0.32    -0.07 1.00
cor(Intercept,sigma_Intercept)     0.23      0.06     0.12     0.34 1.00
cor(BINz,sigma_Intercept)          0.03      0.06    -0.10     0.15 1.00
                               Bulk_ESS Tail_ESS
sd(Intercept)                      2914     2846
sd(BINz)                           2861     2637
sd(sigma_Intercept)                2918     2985
cor(Intercept,BINz)                2840     2619
cor(Intercept,sigma_Intercept)     2767     2943
cor(BINz,sigma_Intercept)          3081     2832
```

```
Population-Level Effects:
                Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept           0.04      0.03    -0.02     0.10 1.00     3085     2775
sigma_Intercept    -0.72      0.02    -0.76    -0.67 1.00     2940     2823
BINz               -0.35      0.01    -0.38    -0.32 1.00     3001     2767
BIN2                0.11      0.01     0.10     0.12 1.00     2795     3070
JUL                -0.15      0.03    -0.20    -0.10 1.00     2673     3020
MIN                -0.05      0.02    -0.09    -0.01 1.00     2988     2868
SEQ                -0.24      0.03    -0.29    -0.19 1.00     2505     2906
SEX                -0.04      0.03    -0.10     0.01 1.00     2858     2929
AGE                -0.06      0.03    -0.12    -0.01 1.00     2984     2840
MAS                -0.05      0.03    -0.11     0.01 1.00     2857     2946
REP                 0.03      0.03    -0.02     0.08 1.00     2942     2846
BINz:JUL            0.04      0.01     0.02     0.07 1.00     2811     2697
BINz:MIN            0.01      0.01    -0.01     0.03 1.00     3102     2870
BINz:SEQ            0.11      0.01     0.08     0.13 1.00     2966     2956
BINz:SEX            0.02      0.01    -0.00     0.05 1.00     2819     2811
BINz:AGE            0.02      0.01    -0.01     0.05 1.00     2910     3033
BINz:MAS           -0.04      0.01    -0.06    -0.01 1.00     2936     2947
BINz:REP            0.00      0.01    -0.02     0.03 1.00     3189     2965
sigma_BINz         -0.13      0.01    -0.15    -0.10 1.00     2895     3029
sigma_BIN2          0.12      0.01     0.09     0.14 1.00     2799     2719
sigma_JUL          -0.05      0.02    -0.10    -0.01 1.00     3108     2943
sigma_MIN          -0.01      0.02    -0.04     0.02 1.00     2966     2723
sigma_SEQ          -0.13      0.02    -0.17    -0.09 1.00     2953     2636
sigma_SEX          -0.01      0.02    -0.05     0.03 1.00     2910     2852
sigma_AGE          -0.03      0.02    -0.07     0.02 1.00     3205     2788
sigma_MAS           0.01      0.02    -0.03     0.06 1.00     3040     2838
sigma_REP           0.02      0.02    -0.02     0.06 1.00     2871     2702

Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

References

Allaire, J. C., & Marsiske, M. (2005). Intraindividual variability may not always indicate vulnerability in elders' cognitive performance. *Psychology and Aging*, *20*(3), 390–401. https://doi.org/10.1037/0882-7974.20.3.390

Araya-Ajoy, Y. G., & Dingemanse, N. J. (2017). Repeatability, heritability, and age-dependence of seasonal plasticity in aggressiveness in a wild passerine bird. *Journal of Animal Ecology*, *86*(2), 227–238. https://doi.org/10.1111/1365-2656.12621

Araya-Ajoy, Y. G., Mathot, K. J., & Dingemanse, N. J. (2015). An approach to estimate short-term, long-term and reaction norm repeatability. *Methods in Ecology and Evolution*, *6*(12), 1462–1473. https://doi.org/10.1111/2041-210X.12430

Archer, J. (1973). Tests for emotionality in rats and mice: A review. *Animal Behaviour*, *21*(2), 205–235. https://doi.org/10.1016/S0003-3472(73)80065-X

Barnett, A. G., van der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology*, *34*(1), 215–220. https://doi.org/10.1093/ije/dyh299

Bell, A. M., Hankison, S. J., & Laskowski, K. L. (2009). The repeatability of behaviour: a meta-analysis. *Animal Behaviour*, *77*(4), 771–783. https://doi.org/10.1016/j.anbehav.2008.12.022

Berberi, I., & Careau, V. (2019). Performance trade-offs in wild mice. *Oecologia*, *191*(1), 11–23. https://doi.org/10.1007/s00442-019-04450-9

Biondi, L. M., Guido, J., Madrid, E., Bó, M. S., & Vassallo, A. I. (2013). The Effect of Age and Sex on Object Exploration and Manipulative Behavior in a Neotropical Raptor, the

Chimango Caracara, Milvago chimango. *Ethology*, *119*(3), 221–232. https://doi.org/10.1111/eth.12056

Biro, P., & Adriaenssens, B. (2013). Predictability as a personality trait: Consistent differences in intraindividual behavioral variation. *American Naturalist*, *182*(5), 621–629. https://doi.org/10.1086/673213

Biro, P., O'Connor, J., Pedini, L., & Gribben, P. E. (2013). Personality and plasticity: Consistent responses within-, but not across-temperature situations in crabs. *Behaviour*, *150*(7), 799–811. https://doi.org/10.1163/1568539X-00003081

Boon, A. K., Reale, D., & Boutin, S. (2008). Personality, habitat use, and their consequences for survival in North American red squirrels Tamiasciurus hudsonicus. *Oikos*, *1*. https://doi.org/10.1111/j.2008.0030-1299.16567.x

Boon, A. K., Réale, D., & Boutin, S. (2007). The interaction between personality, offspring fitness and food abundance in North American red squirrels. *Ecology Letters*, *10*(11), 1094–1104. https://doi.org/10.1111/j.1461-0248.2007.01106.x

Boyer, N., Réale, D., Marmet, J., Pisanu, B., & Chapuis, J. L. (2010). Personality, space use and tick load in an introduced population of Siberian chipmunks Tamias sibiricus. *Journal of Animal Ecology*, *79*(3), 538–547. https://doi.org/10.1111/j.1365-2656.2010.01659.x

Briffa, M. (2013). Plastic proteans: Reduced predictability in the face of predation risk in hermit crabs. *Biology Letters*, *9*(5). https://doi.org/10.1098/rsbl.2013.0592

Buerkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28.

Careau, V., Bininda-Emonds, O. R. P., Ordonez, G., & Garland, T. (2012). Are voluntary wheel running and open-field behavior correlated in mice? different answers from comparative and artificial selection approaches. *Behavior Genetics*, *42*(5), 830–844. https://doi.org/10.1007/s10519-012-9543-0

Careau, V., Mariette, M. M., Crino, O., Buttemer, W. A., & Buchanan, K. L. (2020). Repeatability of behavior and physiology: No impact of reproductive investment. *General and Comparative Endocrinology*, *290*, 113403. https://doi.org/10.1016/j.ygcen.2020.113403

Careau, V., Montiglio, P. O., Garant, D., Pelletier, F., Speakman, J. R., Humphries, M. M., & Réale, D. (2015). Energy expenditure and personality in wild chipmunks. *Behavioral Ecology and Sociobiology*, *69*(4), 653–661. https://doi.org/10.1007/s00265-015-1876-2

Chang, C. C., Teo, H. Y., Norma-Rashid, Y., & Li, D. (2017). Predator personality and prey behavioural predictability jointly determine foraging performance. *Scientific Reports*, *7*, 1–8. https://doi.org/10.1038/srep40734

Cleasby, I. R., Nakagawa, S., & Schielzeth, H. (2015). Quantifying the predictability of behaviour: Statistical approaches for the study of between-individual variation in the within-individual variance. *Methods in Ecology and Evolution*, *6*(1), 27–37. https://doi.org/10.1111/2041-210X.12281

Collins, H. H. (1923). Studies of the pelage phases and of the nature of color variations in mice of the genus Peromyscus. *Journal of Experimental Zoology*, *38*(1), 45–107. https://doi.org/10.1002/jez.1400380103

DeWitt, T. J., Sih, A., & Wilson, D. S. (1998). Costs and limits of phenotypic plasticity. *Trends* *in Ecology and Evolution*. https://doi.org/10.1016/S0169-5347(97)01274-3

Dingemanse, N. J., Both, C., Drent, P. J., Van Oers, K., & Van Noordwijk, A. J. (2002). Repeatability and heritability of exploratory behaviour in great tits from the wild. *Animal Behaviour*, *64*(6), 929–938. https://doi.org/10.1006/anbe.2002.2006

Dingemanse, N. J., & Dochtermann, N. A. (2013). Quantifying individual variation in behaviour: Mixed-effect modelling approaches. *Journal of Animal Ecology*, *82*(1), 39–54. https://doi.org/10.1111/1365-2656.12013

Dingemanse, N. J., Kazem, A. J. N., Réale, D., & Wright, J. (2010). Behavioural reaction norms: animal personality meets individual plasticity. *Trends in Ecology and Evolution*, *25*(2), 81–89. https://doi.org/10.1016/j.tree.2009.07.013

Driver, P. M., & Humphries, D. A. (1970). Protean Defence by Prey Animals. *Oecologia*, *5*(4), 285–302.

Elder, G. A., Ragnauth, A., Dorr, N., Franciosi, S., Schmeidler, J., Haroutunian, V., & Buxbaum, J. D. (2008). Increased locomotor activity in mice lacking the low-density lipoprotein receptor. *Behavioural Brain Research*, *191*(2), 256–265. https://doi.org/10.1016/j.bbr.2008.03.036

Endler, J. A. (1986). *Natural selection in the wild* (N. J. Princeton (ed.)). Princeton University Press.

Feda, D. M., Roberts, A. M., & Roemmich, J. N. (2015). Habituation to a stressor predicts adolescents' adiposity. *Anxiety, Stress and Coping*, *29*(4), 457–462. https://doi.org/10.1080/10615806.2015.1065318

Felleki, M., & Lundeheim, N. (2015). Genetic heteroscedasticity of teat count in pigs. *Journal of Animal Breeding and Genetics*, *132*(5), 392–398. https://doi.org/10.1111/jbg.12134

Fiedler, A. (2019). *Resting and Maximal Metabolic Rates in Wild White-Footed mice (Peromyscus leucopus)*. Univeristy of Ottawa.

Fraser, D. F., Gilliam, J. F., Daley, M. J., Le, A. N., & Skalski, G. T. (2001). Explaining Leptokurtic Movement Distributions: Intrapopulation Variation in Boldness and Exploration. *The American Naturalist*, *158*(2), 124–135. https://doi.org/10.1086/321307

Hall, C. S. (1941). Emotional Behavior in the Rat. *Journal of Comparative Psychology*, *18*(5), 385–403. https://doi.org/10.1037/h0071444

Hall, C. S., & Ballachey, E. L. (1932). A study of the rat's behavior in a field. A contribution to method in comparative psychology. *University of California Publications in Psychology*.

Highcock, L., & Carter, A. J. (2014). Intraindividual variability of boldness is repeatable across contexts in a wild lizard. *PLoS ONE*, *9*(4). https://doi.org/10.1371/journal.pone.0095179

Hill, W. G., & Mulder, H. A. (2010). Genetic analysis of environmental variation. *Genetics Research*, *92*(5–6), 381–395. https://doi.org/10.1017/S0016672310000546

Hoffman, L. (2007). Multilevel Models for Examining Individual Differences in Within-Person Variation and Covariation Over Time. *Multivariate Behavioral Research*, *42*(4), 609–629. https://doi.org/10.1080/00273170701710072

Houslay, T. M., Vierbuchen, M., Grimmer, A. J., Young, A. J., & Wilson, A. J. (2018). Testing the stability of behavioural coping style across stress contexts in the Trinidadian guppy. *Functional Ecology*, *32*(2), 424–438. https://doi.org/10.1111/1365-2435.12981

Jennings, D. J., Hayden, T. J., & Gammell, M. P. (2013). Personality and predictability in fallow deer fighting behaviour: The relationship with mating success. *Animal Behaviour*, *86*(5), 1041–1047. https://doi.org/10.1016/j.anbehav.2013.09.009

Jolles, J. W., Briggs, H. D., Araya-Ajoy, Y. G., & Boogert, N. J. (2019). Personality, plasticity and predictability in sticklebacks: bold fish are less plastic and more predictable than shy fish. *Animal Behaviour*, *154*, 193–202. https://doi.org/10.1016/j.anbehav.2019.06.022

Jones, K. A., Jackson, A. L., & Ruxton, G. D. (2011). Prey jitters; Protean behaviour in grouped prey. *Behavioral Ecology*, *22*(4), 831–836. https://doi.org/10.1093/beheco/arr062

Kelly, C., & Price, T. D. (2005). Correcting for regression to the mean in behavior and ecology. *American Naturalist*, *166*(6), 700–707. https://doi.org/10.1086/497402

Lee, A. H., Wang, K., Yau, K. K. W., & Somerford, P. J. (2003). Truncated negative binomial mixed regression modelling of ischaemic stroke hospitalizations. *Statistics in Medicine*, *22*, 1129–1139. https://doi.org/10.1002/sim.1419

Martin, J. G. A., Nussey, D. H., Wilson, A. J., & Réale, D. (2011). Measuring individual differences in reaction norms in field and experimental studies: A power analysis of random regression models. *Methods in Ecology and Evolution*, *2*(4), 362–374. https://doi.org/10.1111/j.2041-210X.2010.00084.x

Martin, J. G. A., Pirotta, E., Petelle, M. B., & Blumstein, D. T. (2017). Genetic basis of between-individual and within-individual variance of docility. *Journal of Evolutionary Biology*, *30*(4), 796–805. https://doi.org/10.1111/jeb.13048

Martin, J. G. A., & Réale, D. (2008). Temperament, risk assessment and habituation to novelty in eastern chipmunks, Tamias striatus. *Animal Behaviour*, *75*(1), 309–318. https://doi.org/10.1016/j.anbehav.2007.05.026

Miller, R., Bugnyar, T., Pölzl, K., & Schwab, C. (2015). Differences in exploration behaviour in common ravens and carrion crows during development and across social context. *Behavioral Ecology and Sociobiology*, *69*(7), 1209–1220. https://doi.org/10.1007/s00265-015-1935-8

Mitchell, D. J., Fanson, B. G., Beckmann, C., & Biro, P. A. (2016). Towards powerful experimental and statistical approaches to study intraindividual variability in labile traits. *Royal Society Open Science*, *3*(10). https://doi.org/10.1098/rsos.160352

Montiglio, P. O., Garant, D., Thomas, D., & Réale, D. (2010). Individual variation in temporal activity patterns in open-field tests. *Animal Behaviour*, *80*(5), 905–912. https://doi.org/10.1016/j.anbehav.2010.08.014

Mulder, H. A., Gienapp, P., & Visser, M. E. (2016). Genetic variation in variability: Phenotypic variability of fledging weight and its evolution in a songbird population. *Evolution; International Journal of Organic Evolution*, *70*(9), 2004–2016. https://doi.org/10.1111/evo.13008

Mulder, H. A., Hill, W. G., Vereijken, A., & Veerkamp, R. F. (2009). Estimation of genetic variation in residual variance in female and male broiler chickens. *Animal*, *3*(12), 1673–1680. https://doi.org/10.1017/S1751731109990668

Mulder, H. A., Rönnegård, L., Fikse, W. F., Veerkamp, R. F., & Strandberg, E. (2013). Estimation of genetic variance for macro- and micro-environmental sensitivity using double hierarchical generalized linear models. *Genetics Selection Evolution*, *45*(1), 1–14.

https://doi.org/10.1186/1297-9686-45-23

Natarajan, D., De Vries, H., Saaltink, D. J., De Boer, S. F., & Koolhaas, J. M. (2009). Delineation of violence from functional aggression in mice: An ethological approach. *Behavior Genetics*, *39*(1), 73–90. https://doi.org/10.1007/s10519-008-9230-3

Nesselroade, J. R., Stigler, S. M., & Baltes, P. B. (1980). Regression toward the mean and the study of change. *Psychological Bulletin*, *88*(3), 622–637. https://doi.org/10.1037/0033-2909.88.3.622

Neves, H. H. R., Carvalheiro, R., & Queiroz, S. A. (2012). Genetic and environmental heterogeneity of residual variance of weight traits in Nellore beef cattle. *Genetics Selection Evolution*, *44*(1), 1–12. https://doi.org/10.1186/1297-9686-44-19

Newar, S. L., & Careau, V. (2018). The fast and the curious: locomotor performance and exploratory behaviour in eastern chipmunks. *Behavioral Ecology and Sociobiology*, *72*(2). https://doi.org/10.1007/s00265-018-2445-2

Noftle, E. E., & Fleeson, W. (2010). Age Differences in Big Five Behavior Averages and Variabilities Across the Adult Life Span: Moving Beyond Retrospective, Global Summary Accounts of Personality. *Psychology and Aging*, *25*(1), 95–107. https://doi.org/10.1037/a0018199

Nussey, D. H., Postma, E., Gienapp, P., & Visser, M. E. (2005). Selection on heritable phenotypic plasticity in a wild bird population. *Science*, *310*(5746), 304–306. https://doi.org/10.1126/science.1117004

Rankin, C. H., Abrams, T., Barry, R. J., Bhatnagar, S., Clayton, D. F., Colombo, J., Coppola, G., Geyer, M. A., Glanzman, D. L., Marsland, S., McSweeney, F. K., Wilson, D. A., Wu, C. F., & Thompson, R. F. (2009). Habituation revisited: An updated and revised description of the behavioral characteristics of habituation. *Neurobiology of Learning and Memory*, *92*(2), 135–138. https://doi.org/10.1016/j.nlm.2008.09.012

Ray, J., & Hansen, S. (2005). Temperamental development in the rat: The first year. *Developmental Psychobiology*, *47*(2), 136–144. https://doi.org/10.1002/dev.20080

Reader, S. M. (2015). Causes of Individual Differences in Animal Exploration and Search. *Topics in Cognitive Science*, *7*(3), 451–468. https://doi.org/10.1111/tops.12148

Réale, D., Reader, S. M., Sol, D., McDougall, P. T., & Dingemanse, N. J. (2007). Integrating animal temperament within ecology and evolution. *Biological Reviews*, *82*(2), 291–318. https://doi.org/10.1111/j.1469-185X.2007.00010.x

Richardson, G., Dickinson, P., Burman, O. H. P., & Pike, T. W. (2018). Unpredictable movement as an anti-predator strategy. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1885). https://doi.org/10.1098/rspb.2018.1112

Rödel, H. G., Zapka, M., Talke, S., Kornatz, T., Bruchner, B., & Hedler, C. (2015). Survival costs of fast exploration during juvenile life in a small mammal. *Behavioral Ecology and Sociobiology*, *69*(2), 205–217. https://doi.org/10.1007/s00265-014-1833-5

Rönnegård, L., Felleki, M., Fikse, F., Mulder, H. A., & Strandberg, E. (2010). Genetic heterogeneity of residual variance - Estimation of variance components using double hierarchical generalized linear models. *Genetics Selection Evolution*, *42*(1), 1–10. https://doi.org/10.1186/1297-9686-42-8

Rosengren, K. S., & Braswell, G. S. (2001). Variability in Children's Reasoning. *Advances In Child Development and Behaviour*, *28*, 1–40.

Schaeffer, L. R. (2004). Application of random regression models in animal breeding. *Livestock Production Science*, *86*(1–3), 35–45. https://doi.org/10.1016/S0301-6226(03)00151-9

Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution*, *1*(2), 103–113. https://doi.org/10.1111/j.2041-210x.2010.00012.x

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2009). On the Relation of Mean Reaction Time and Intraindividual Reaction Time Variability. *Psychology and Aging*, *24*(4), 841–857. https://doi.org/10.1037/a0017799

Schuster, A. C., Carl, T., & Foerster, K. (2017). Repeatability and consistency of individual behaviour in juvenile and adult eurasian harvest mice. *Science of Nature*, *104*(3–4), 1–14. https://doi.org/10.1007/s00114-017-1430-3

Smith, G. M., Saveliev, A. A., Walker, N., Ieno, E. N., & Zuur, A. F. (2009). Zero-Truncated and Zero-Inflated Models for Count Data. In *Mixed Effects Models and Extensions in Ecology with R* (Vol. 1, pp. 261–293). Springer New York. https://doi.org/10.1007/978-0-387-87458-6

Sonnweber, R., Araya-Ajoy, Y. G., Behringer, V., Deschner, T., Tkaczynski, P., Fedurek, P., Preis, A., Samuni, L., Zommers, Z., Gomes, C., Zuberbühler, K., Wittig, R. M., & Crockford, C. (2018). Circadian rhythms of urinary cortisol levels vary between individuals in wild male chimpanzees: A reaction norm approach. *Frontiers in Ecology and Evolution*, *6*(JUN), 1–11. https://doi.org/10.3389/fevo.2018.00085

Stamps, J. A., Briffa, M., & Biro, P. A. (2012). Unpredictable animals: Individual differences in intraindividual variability (IIV). *Animal Behaviour*, *83*(6), 1325–1334. https://doi.org/10.1016/j.anbehav.2012.02.017

Stamps, J. A., & Krishnan, V. V. (2014). Combining information from ancestors and personal experiences to predict individual differences in developmental trajectories. *American Naturalist*, *184*(5), 647–657. https://doi.org/10.1086/678116

Stigler, S. M. (1997). Regression towards the mean, historically considered. *Statistical Methods in Medical Research*, *6*(2), 103–114. https://doi.org/10.1177/096228029700600202

Szechtman, H., Talangbayan, H., Canaran, G., Dai, H., & Eilam, D. (1994). Dynamics of behavioral sensitization induced by the dopamine agonist quinpirole and a proposed central energy control mechanism. *Psychopharmacology*, *115*(1–2), 95–104. https://doi.org/10.1007/BF02244757

Sztepanacz, J. L., McGuigan, K., & Blows, M. W. (2017). Heritable micro-environmental variance covaries with fitness in an outbred population of Drosophila serrata. *Genetics*, *206*(4), 2185–2198. https://doi.org/10.1534/genetics.116.199075

Tonsor, S. J., Elnaccash, T. W., & Scheiner, S. M. (2013). Developmental instability is genetically correlated with phenotypic plasticity, constraining heritability, and fitness. *Evolution*. https://doi.org/10.1111/evo.12175

Van Overveld, T., & Matthysen, E. (2010). Personality predicts spatial responses to food manipulations in free-ranging great tits (Parus major). *Biology Letters*, *6*(2), 187–190. https://doi.org/10.1098/rsbl.2009.0764

Vanden Broecke, B., Borremans, B., Mariën, J., Makundi, R. H., Massawe, A. W., Leirs, H., & Hughes, N. K. (2018). Does exploratory behavior or activity in a wild mouse explain susceptibility to virus infection? *Current Zoology*, *64*(5), 585–592. https://doi.org/10.1093/cz/zox053

Vessey, S. H., & Vessey, K. B. (2007). Linking behavior, life history and food supply with the population dynamics of white-footed mice (Peromyscus leucopus). *Integrative Zoology*, *2*(3), 123–130. https://doi.org/10.1111/j.1749-4877.2007.00053.x

Walsh, R. N., & Cummins, R. A. (1976). The open-field test: A critical review. *Psychological Bulletin*, *83*(3), 482–504. https://doi.org/10.1037/0033-2909.83.3.482

Warren, E. W., & Callaghan, S. (1975). Individual differences in response to an open field test by the guppy—Poecilia reticulata (Peters). *Journal of Fish Biology*, *7*(1), 105–113. https://doi.org/10.1111/j.1095-8649.1975.tb04580.x

Westneat, D. F., Hatch, M. I., Wetzel, D. P., & Ensminger, A. L. (2011). Individual Variation in Parental Care Reaction Norms: Integration of Personality and Plasticity. *The American Naturalist*, *178*(5), 652–667. https://doi.org/10.1086/662173

Westneat, D. F., Schofield, M., & Wright, J. (2013). Parental behavior exhibits among-individual variance, plasticity, and heterogeneous residual variance. *Behavioral Ecology*, *24*(3), 598–604. https://doi.org/10.1093/beheco/ars207

Westneat, D. F., Wright, J., & Dingemanse, N. J. (2015). The biology hidden inside residual within-individual phenotypic variation. *Biological Reviews*, *90*(3), 729–743. https://doi.org/10.1111/brv.12131

Williams, B. R., Hultsch, D. F., Strauss, E. H., Hunter, M. A., & Tannock, R. (2005). Inconsistency in reaction time across the life span. *Neuropsychology*, *19*(1), 88–96. https://doi.org/10.1037/0894-4105.19.1.88

Wolf, M., Van Doorn, G. S., & Weissing, F. J. (2011). On the coevolution of social responsiveness and behavioural consistency. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1704), 440–448. https://doi.org/10.1098/rspb.2010.1051

Wolff, J. O. (1985). The effects of density, food, and interspecific interference on home range size in Peromyscus leucopus and Peromyscus maniculatus. *Canadian Journal of Zoology*, *63*(11), 2657–2662. https://doi.org/https://doi.org/10.1139/z85-397

Wolff, J. O. (1996). Coexistence of white-footed mice and deer mice may be mediated by fluctuating environmental conditions. *Oecologia*, *108*(3), 529–533. https://doi.org/10.1007/BF00333730

Xia, X., Millar, J. S., & Na, C. (1991). Genetic evidence of promiscuity in Peromyscus leucopus. *Behavioral Ecology and Sociobiology*, 171–178.

Tables and Figures

Table 1. Sources of variation in 909 behavioural reaction norms of 436 wild white-footed mice (*Peromyscus leucopus*) during a 10-min open-field test, extracted from a doubly hierarchical generalized linear mixed model. Distance moved during each minute of the test was analyzed as function of test minute, test minute^2, Julian day, time of day, test sequence, sex, age, body mass, and reproductive status, fitted as main effects to both the intercept (personality) and standard deviation (unpredictability) of the reaction norms. Main effects were also included in interaction with test minute to test for their effect on the slope of the behavioural reaction norms (plasticity). Shown are estimates of each fixed effect with their standard error (SE), and their 95% upper and lower limits of credible intervals (see Table 2 for random effects). Bold values indicate significance.

Level	Source	Estimate ± SE	Credible Interval	
			Lower	Upper
Intercept (Personality)	Intercept	0.041 ± 0.030	-0.019	0.102
	Test minute	**-0.349 ± 0.015**	**-0.378**	**-0.320**
	Test minute^2	**0.107 ± 0.006**	**0.096**	**0.119**
	Julian day	**-0.148 ± 0.026**	**-0.199**	**-0.096**
	Time of day	**-0.046 ± 0.020**	**-0.086**	**-0.007**
	Test sequence	**-0.242 ± 0.025**	**-0.289**	**-0.193**
	Sex$_{[male]}$	-0.045 ± 0.027	-0.097	0.007
	Age$_{[juvenile]}$	**-0.065 ± 0.027**	**-0.118**	**-0.011**
	Body mass	-0.051 ± 0.028	-0.107	0.005
	Reproductive status$_{[active]}$	0.029 ± 0.025	-0.021	0.077
Slope (Plasticity)	Test minute × Julian day	**0.044 ± 0.013**	**0.020**	**0.069**
	Test minute × Time of day	0.006 ± 0.010	-0.014	0.026
	Test minute × Test sequence	**0.106 ± 0.012**	**0.081**	**0.130**
	Test minute × Sex	0.024 ± 0.013	0.000	0.049
	Test minute × Age	0.023 ± 0.014	-0.005	0.051
	Test minute × Body mass	**-0.036 ± 0.014**	**-0.064**	**-0.009**
	Test minute × Reproductive status	0.002 ± 0.013	-0.023	0.027
Standard deviation (Unpredictability)	Intercept	**-0.717 ± 0.023**	**-0.764**	**-0.671**
	Test minute	**-0.127 ± 0.011**	**-0.149**	**-0.104**
	Test minute^2	**0.116 ± 0.012**	**0.093**	**0.140**
	Julian day	**-0.054 ± 0.021**	**-0.096**	**-0.013**
	Time of day	-0.009 ± 0.017	-0.043	0.023
	Test sequence	**-0.133 ± 0.021**	**-0.173**	**-0.093**
	Sex$_{[male]}$	-0.012 ± 0.020	-0.051	0.028
	Age$_{[juvenile]}$	-0.029 ± 0.023	-0.074	0.016
	Body mass	0.015 ± 0.023	-0.029	0.059
	Reproductive status$_{[active]}$	0.025 ± 0.021	-0.018	0.065

Table 2. Random effect variance components from a doubly hierarchical generalized linear mixed model of distance moved in a 10-min open-field test. Individual identity was included as a random effect to quantify the among-individual variance in intercept (V_{int}), slope (V_{slp}), and standard deviation (V_{sd}). A random effect grouping 1-min observations made within a given test ("series") captured within-individual variance in intercept, slope, and standard deviation of the reaction norms. Correlations (r) among reaction norm parameters are presented at the among- and within-individual levels. Repeatability of the intercept and slope of the reaction norm were calculated as $V_{\text{int among}} / (V_{\text{int among}} + V_{\text{int within}})$ and $V_{\text{slp among}} / (V_{\text{slp among}} + V_{\text{slp within}})$. Repeatability of unpredictability (standard deviation of the reaction norm residuals) was calculated as: $V_{\text{sd among}} / (V_{\text{sd among}} + V_{\text{sd within}})$. All values are reported as estimates with the 95% upper and lower limits of credible intervals. Bold values indicate significance.

	Estimate	Credible Interval	
		Lower	Upper
(a) Among individuals:			
V_{int}	0.130	0.085	0.185
V_{slp}	0.016	0.007	0.025
V_{sd}	0.037	0.018	0.057
$r_{\text{int.slp}}$	**-0.505**	**-0.754**	**-0.182**
$r_{\text{sd.int}}$	**0.924**	**0.643**	**0.998**
$r_{\text{sd.slp}}$	**-0.513**	**-0.793**	**-0.115**
(b) Within individuals (among series):			
V_{int}	0.217	0.187	0.257
V_{slp}	0.046	0.036	0.057
V_{sd}	0.157	0.128	0.184
$r_{\text{int.slp}}$	**-0.192**	**-0.319**	**-0.073**
$r_{\text{sd.int}}$	**0.256**	**0.125**	**0.341**
$r_{\text{sd.slp}}$	0.023	-0.104	0.146
(c) Repeatability:			
Intercept	**0.387**	**0.267**	**0.489**
Slope	**0.230**	**0.122**	**0.382**
Variance	**0.198**	**0.104**	**0.285**

Figure Captions

Figure 1. Overhead view of the circular open-field arena (radius=70cm) used for behavioural testing. Shown is one individual white-footed mouse (*Peromyscus leucopus*) in the centre of the arena during a 10-min open-field test. The yellow dotted line represents the individual's past movement path, while the red dot shows the tracked centre point of the individual using Ethovision^{XT} tracking software.

Figure 2. Distance moved (metres) in a 10-min open-field test as function of test minute in 436 white-footed mice (*Peromyscus leucopus*) tested a total of 909 times. Shown are two representative individuals with either high (green line, triangle symbol) or low (blue line, square symbol) residual variance (unpredictability). Grey circles represent all observations, black circles show average activity for each minute (±SD).

Figure 3. Among-individual (left panels) and within-individual (right panels) correlations between reaction norm parameters in 436 wild white-footed mice (*Peromyscus leucopus*), including the correlations between A-B) reaction norm intercept (personality) and slope (plasticity), C-D) reaction norm intercept (personality) and residual variance (unpredictability), and E-F) reaction norm slope (plasticity) and residual variance (unpredictability). Individual deviations from the population mean are illustrated using best linear unbiased predictors (BLUP) with their 95% credible intervals (CI) associated with the random effect of individual identity. BLUPs associated to the random effect of test series were extracted to illustrate within-individual deviations from individual means.

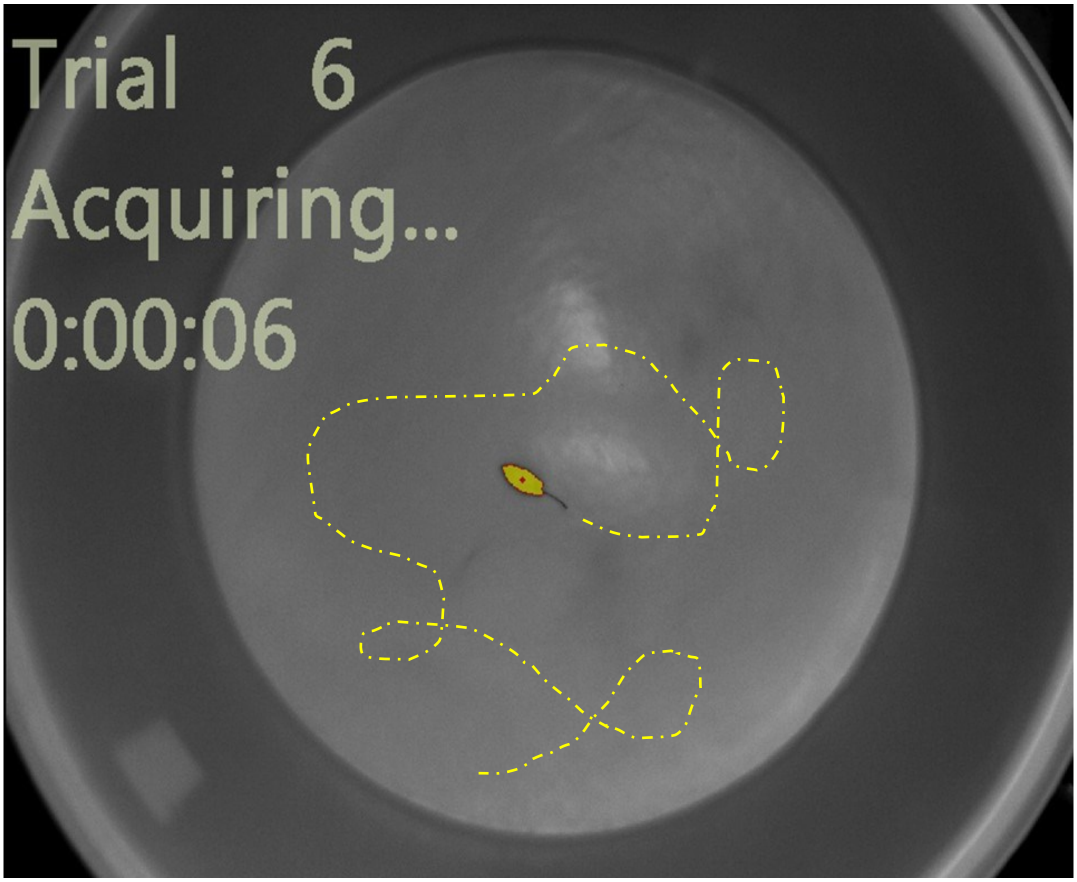
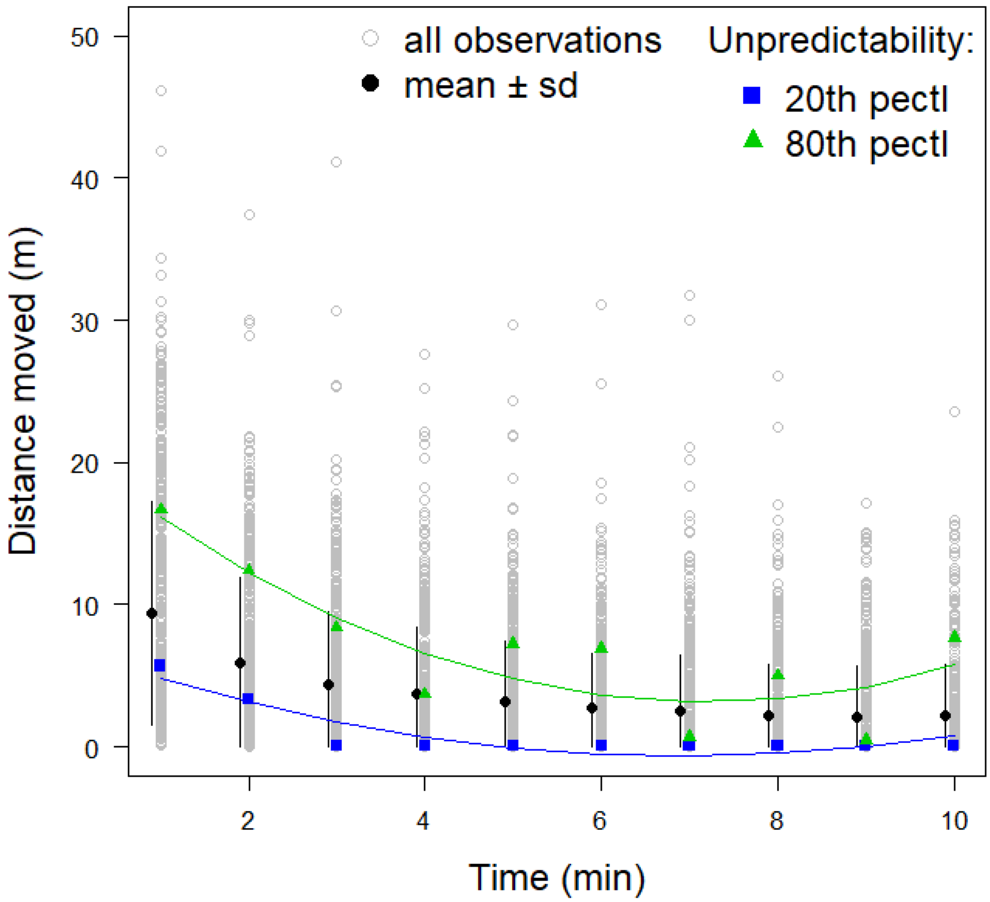


Figure 1



all observations
mean ± sd
Unpredictability:
20th pectl
80th pectl
Distance moved (m)
0
10
20
30
40
50
2
4
6
8
10
Time (min)

Figure 2

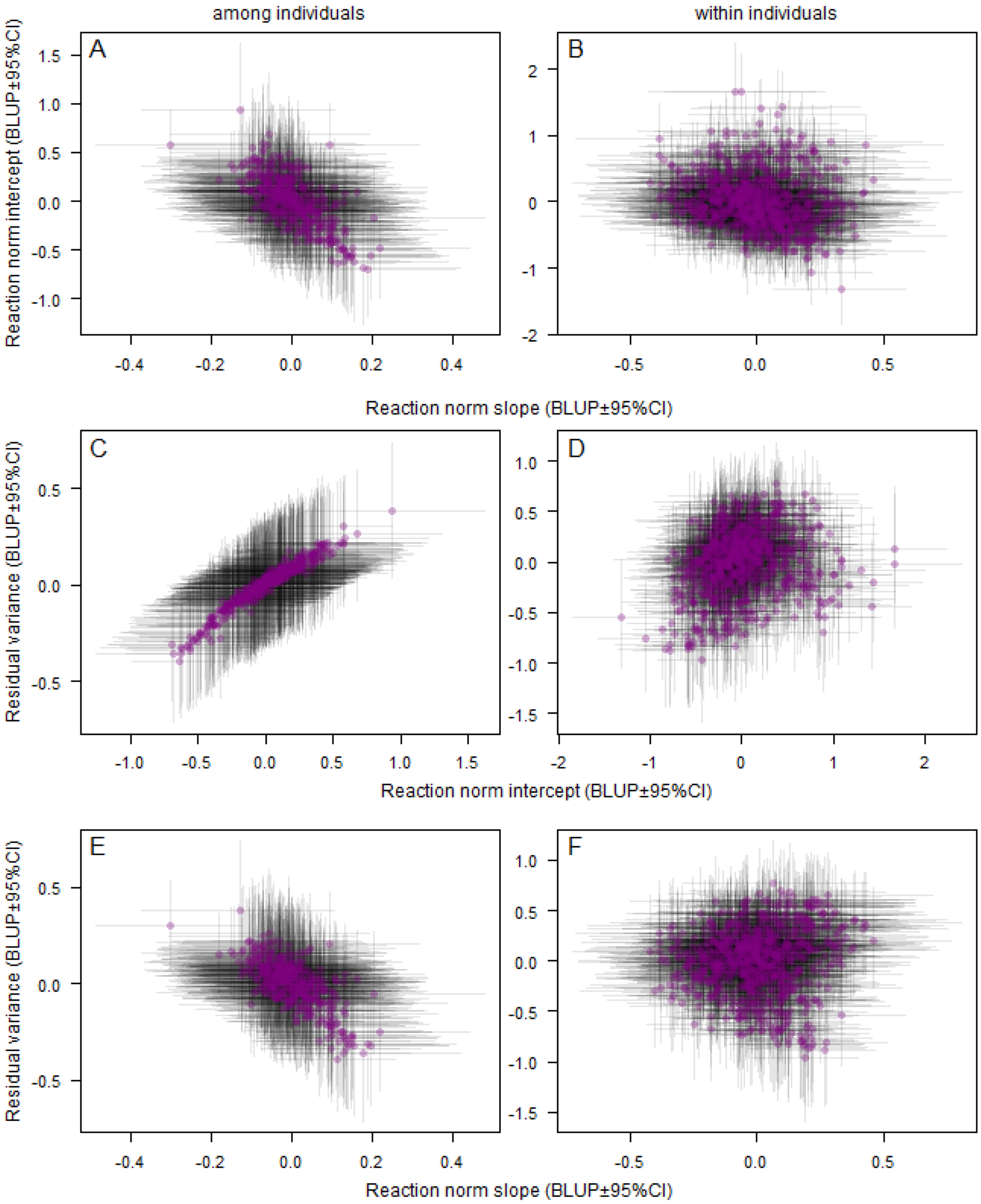

among individuals
within individuals
A
B
C
D
E
F
Reaction norm intercept (BLUP±95%CI)
Reaction norm slope (BLUP±95%CI)
Residual variance (BLUP±95%CI)
Reaction norm intercept (BLUP±95%CI)
Residual variance (BLUP±95%CI)
Reaction norm slope (BLUP±95%CI)
1.5
1.0
0.5
0.0
-0.5
-1.0
-0.4
-0.2
0.0
0.2
0.4
2
1
0
-1
-2
-0.5
0.0
0.5
-1.0
-0.5
0.0
0.5
1.0
1.5
-1.5
-2
-1
0
1
2

Figure 3

Appendix

Table S1. Random effect variance components from a doubly hierarchical generalized linear mixed model of distance moved in a 5-min open-field test. Individual identity was included as a random effect to quantify the among-individual variance in intercept (V_{int}), slope (V_{slp}), and standard deviation (V_{sd}). A random effect grouping 1-min observations made within a given test ("series") captured within-individual variance in intercept, slope, and standard deviation of the reaction norms. Correlations (r) among reaction norm parameters are presented at the among- and within-individual levels. Repeatability of the intercept and slope of the reaction norm were calculated as $V_{\mathrm{int\ among}} / (V_{\mathrm{int\ among}} + V_{\mathrm{int\ within}})$ and $V_{\mathrm{slp\ among}} / (V_{\mathrm{slp\ among}} + V_{\mathrm{slp\ within}})$. Repeatability of unpredictability (standard deviation of the reaction norm residuals) was calculated as: $V_{\mathrm{sd\ among}} / (V_{\mathrm{sd\ among}} + V_{\mathrm{sd\ within}})$. All values are reported as estimates with the 95% upper and lower limits of credible intervals. Bold values indicate significance.

	Estimate	Credible Interval	
		Lower	Upper
(*a*) Among individuals:			
V_{int}	0.185	0.117	0.241
V_{slp}	0.021	0.012	0.031
V_{sd}	0.014	0.000	0.045
$r_{\mathrm{int.slp}}$	-0.204	-0.465	0.065
$r_{\mathrm{sd.int}}$	**0.656**	**0.098**	**0.964**
$r_{\mathrm{sd.slp}}$	-0.368	-0.840	0.164
(*b*) Within individuals (among series):			
V_{int}	0.270	0.235	0.326
V_{slp}	0.025	0.015	0.034
V_{sd}	0.139	0.098	0.179
$r_{\mathrm{int.slp}}$	**-0.312**	**-0.476**	**-0.145**
$r_{\mathrm{sd.int}}$	0.026	-0.134	0.177
$r_{\mathrm{sd.slp}}$	0.089	-0.165	0.278
(c) Repeatability:			
Intercept	**0.381**	**0.284**	**0.501**
Slope	**0.480**	**0.299**	**0.647**
Variance	0.099	0.000	0.271