Title
The Causes and Maintenance of Personality in Yellow-bellied Marmots (Marmota flaviventris)

Author
Petelle, Matthew Brian

Publication Date
2014

UNIVERSITY OF CALIFORNIA

Los Angeles

The Causes and Maintenance of Personality in Yellow-bellied Marmots (*Marmota flaviventris*)

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Biology

by

Matthew Brian Petelle

2014

2014

ABSTRACT OF THE DISSERTATION

The Causes and Consequences of Personality in Yellow-bellied Marmots (*Marmota flaviventris*)

by

Matthew Brian Petelle
Doctor of Philosophy in Biology
University of California, Los Angeles, 2014
Professor Daniel T. Blumstein, Chair

Animal personality, or consistent individual differences in behavior, is wide spread across taxa, and is now being linked to ecology and evolutionary dynamics. Despite interest in the ecological and evolutionary consequences of personality, few studies have used a Tinbergian approach to understanding the causes and maintenance of personality. Furthermore, there is a large amount of variation within personality traits, and as evolutionary biologists, we are keenly interested in how variation is caused, develops, and is maintained within a population. My dissertation uses Tinbergian principles to try to explain personality using yellow-bellied marmots (*Marmota flaviventris*) as a study system. First, I explore the methodology that describes personality traits by testing whether unacquainted raters could reliably assess subjects using subjective ratings. I found that raters could reliably measure subjects, and some of these measures were valid when compared to behavioral codings. I then focus on the development of personality and found that docility and boldness do not follow the same ontogenetic path. These traits become repeatable at different life stages, and this may reflect differences in stage-specific life history strategies. I also found that boldness and docility do not form a behavioral syndrome, and that this is most likely due to the differences in development. My next chapter focuses on the causes and maintenance of personality. I test three major theoretical hypotheses – growth-mortality tradeoffs, residual reproductive value, and state-dependent safety – and found no evidence for any. I did, however, find that different environmental variables differentially influence the same personality traits across contexts suggesting that selection can influence the same personality trait through different variables depending on the context. Finally, I explore the quantitative genetics of personality. To fully understand the evolution of personality, we need to know the heritability and correlations underlying these traits. I found low heritability in most personality traits with some correlations. This dissertation shows, that in marmots, personality is heavily influenced by environment and that personality is linked to life history strategy.

The dissertation of Matthew Brian Petelle is approved.

Lynn Fairbanks

Peter Nonacs

Daniel T. Blumstein, Committee Chair

University of California, Los Angeles

2014

TABLE OF CONTENTS

LIST OF TABLES vi

LIST OF FIGURES xi

ACKNOWLEDGMENTS xii

VITA xiv

CHAPTER 1. General Introduction 1

- Study System and General Methods 2
- Validating Personality Measurements 5
- Development of Personality 6
- Causes and Maintenance of Personality 8
- Quantitative Genetics of Personality 9
- References 11

CHAPTER 2. A Critical Evaluation of Subjective Ratings: Unacquainted Observers Can Reliably Assess Certain Personality Traits 15

- Introduction 15
- Methods 16
- Results 18
- Discussion 19
- References 21

CHAPTER 3. Development of Docility and Boldness in Yellow-bellied Marmots 23

- Introduction 23
- Methods 24
- Results 26
- Discussion 27
- References 29

CHAPTER 4. Maintenance of Risky Personality Traits in Yellow-bellied Marmots 31

- Introduction 32
- Methods 34
- Results 39
- Discussion 40
- Supplementary Material 51
- References 56

CHAPTER 5. Heritability and Genetic Correlations of Personality Traits in Yellow-bellied Marmots 60

Introduction....61
Methods....63
Results....68
Discussion....70
Supplementary Material....78
References....86

LIST OF TABLES

Table 1. Intra-class correlation coefficients of adjectives used to describe yellow-bellied marmots in open-field and mirror image stimulation tests 17

Table 2. Summary of principle component analysis for open-field (OF) 18

Table 3. Summary of principle component analysis for mirror image stimulation 19

Table 1. Fixed effects explaining variation in the univariate model of boldness (quantified as the negative of flight initiation distance) for three age-classes in yellow-bellied marmots. Significant effects are in bold 26

Table 2. Fixed effects explaining variation in the univariate model of trapping behaviors, a measure of docility, for three age-classes in yellow-bellied marmots. Significant effects are in bold 26

Table 3. Variance, ratio, and significance of random effects using log-likelihood ratio tests (LRT) for both docility and boldness univariate calculated from models for juvenile, yearling, and adult yellow-bellied marmots. Total number of observations, number of individuals and mean of the traits are also reported. All the LRTs have only 1 degree of freedom 27

Table 4. Repeatability (on diagonal), correlations (below diagonal), and pairwise comparison of repeatability (above diagonal) for docility and boldness for the different age classes (juvenile, yearling, adult) in yellow-bellied marmots. Estimates were obtained from trivariate models for each behaviour considering each age class as a different trait. Estimates of repeatability are slightly different from univariate models (Table 1) because only significant fixed effects were used in these models to avoid overparametrization 27

Table 4-1. Variance components, variance ratio, log-likelihood ratio tests (LRT) and significance of random effects with and without growth, age, and mass for docility (in the trap) and activity and exploration during open field (OF) and mirror image stimulation (MIS) tests. Table 1a reports information for models without growth rate, age class, and mass as fixed effects, while Table 1b describes the models with them. Activity and exploration in OF and MIS were estimated from principal component analyses (factors 1 and 2 for OF, factor 1 and 3 for MIS respectively) and have a mean of 0 and a standard deviation of one. Total number of individuals, trials, mean of the traits, and standard deviation are also reported. We added four and square root transformed Exploration (MIS 3) to normalize residuals 47

Table 4-2. Estimate, standard error, and p-values for growth, age (as a quadratic function), and mass for all docility and activity and exploration in both OF and MIS. We provide both between and within individual estimates for growth and mass....................49

Supplementary Table 1. Principle component analysis of behaviours scored in both open field (OF) and mirror image stimulation (MIS) tests. Loadings with variance explained by each component and total overall variance for components with an eigenvalue greater than one. We considered a variable significantly loaded onto a component if it had a value greater than |0.500|. Defecate, urinate, and immediately ran out of the trap were scored as happening with a dichotomous value (0/1) 51

Supplementary Table 2. Coefficients and standard errors for fixed effects of models for docility and activity and exploration in both open field (OF) and mirror image stimulation (MIS) tests with growth, age, and mass 54

Table 5-1. Heritability (h2 = VA/VP), permanent environment effects (PE = VPE/VP), maternal effects (m2 = VME/VP), year effects (YE = VYE/VP), residual effects, and repeatability (VA + VPE/VP) for docility, activity and exploration in both contexts, and sociability. All effects are given with the equivalent of 95% confidence intervals in parentheses 76

Supplementary Table 1. Principle component analysis of open field (OF) and mirror image stimulation (MIS) tests. Components were Varimax rotated. Variance for each component and total variance explained are at the bottom of the table. We considered any variable over |0.500| as being significantly loaded onto that component. Significant loadings are bolded 78

Supplementary Table 2. Traits, random effects, deviance information criteria (DIC), and delta (Δ) DIC for docility, activity and exploration in OF and MIS tests, and sociability. Random effects were removed from the full model (Individual + Maternal effect + Year) one at a time and delta DIC was calculated. Random effects with delta DIC greater than 2 were considered to make the model significantly better..80

Supplementary Table 3. Phenotypic variances, covariances, and correlations (P-matrix) of yellow-bellied marmot personality traits. Variances are present on the diagonal, the upper triangle contains the correlations, and the bottom triangle the covariances. Correlations and covariances were considered significant if they were different than 0 (based on the Bayesian equivalent of a 95% confidence interval). Significant values are in bold. Non-significant correlations over |0.400| have an asterisk 82

Supplementary Table 4. Additive genetic variances, covariances, and correlations (G-matrix) of yellow-bellied marmot personality traits with the equivalent 95% confidence interval in parentheses. Variances are present on the diagonal, the upper triangle contains the correlations, and the bottom triangle the covariances. Correlations and covariances were considered significant if they were different than 0 (based on the Bayesian equivalent of a 95% confidence interval). Significant values are in bold. Non-significant correlations over |0.400| have an Asterisk 83

Supplementary Table 5. Permanent environment variances, covariances, and correlations of yellow-bellied marmot personality traits. Variances are present on the diagonal, the upper triangle contains the correlations, and the bottom triangle the covariances. Correlations and covariances were considered significant if they were different than 0 (based on the Bayesian equivalent of a 95% confidence interval). Significant values are in bold. Non-significant correlations over |0.400| have an asterisk 84

Supplementary Table 6. Maternal effect variances, covariances, and correlations of yellow-bellied marmot personality traits. Variances are present on the diagonal, the upper triangle contains the correlations, and the bottom triangle the covariances. Correlations and covariances were considered significant if the were different than 0 (based on the Bayesian equivalent of a 95% confidence interval). Significant values are in bold. Non-significant correlations over |0.400| have an asterisk 85

LIST OF FIGURES

Figure 1. Relationship between juvenile/yearling, juvenile/adult, and yearling/adult personality. The top panel represents boldness (negative FID). The bottom panel represents docility (trapping behaviours). Best linear unbiased predictors (BLUPs) from trivariate models of boldness and docility were used for illustration purposes only. N represents the number of individuals for which data were available for both traits 28

Figure 2. Relationship between docility and boldness for a) juveniles, b) yearlings, and c) adults. Best linear unbiased predictors (BLUPs) from bivariate models of docility and boldness were used for illustration purposes only. N represents the number of individual for which data were available for both traits 29

Figure 4-1. Relationship between personality traits docility (panels a-c), activity (d-i) and exploration (j-o) and growth (difference in grams between mass on the 15th of June and 15th of August), age (in years), and mass at capture (in grams). Activity and exploration were estimated from principal components analyses of behaviours in both open field (OF) and mirror image stimulation (MIS) tests (factors 1 and 2 for OF, factor 1 and 3 for MIS respectively) 50

ACKNOWLEDGMENTS

I want to first thank my advisor Dan Blumstein. Throughout my time here at UCLA he has always been enthusiastic about my work and has pushed me to do better. I have learned so much from him, and this work would not have been possible without him. I would also like to thank my committee, Lynn Fairbanks and Peter Nonacs. Their comments have been invaluable.

I would also like to thank my friend and mentor Julien Martin. His help and friendship over the last few years cannot be measured. I also want to thank Raquel Moncús for always being there when I needed a friend. Thank you for always being free and easy with your kindness and friendship. My lab mates and fellow graduate students have also been an important and vital part of my work and experience at UCLA. I especially want to thank Zac Schakner and Amanda Lea for being great friends. Zac, thanks for pushing me at everything and forcing me to be proactive. Amanda, thank you for laughing with me, cooking delicious bread, and being a good friend. I would also like to thank Nicole Munoz for being a great office mate and confidant. Thanks go to all the other Blumstein lab members for their help; Tina Wey, Tiffany Armenta, Megan Owen, Adriana Maldonado, Jenn Smith, Alex Kirschel, Janice Daniel, Oded Berger-Tal, and Eric Abelson.

So many others have helped me throughout my time at UCLA – I would like thanks all the field assistants especially Leon Chong, Casey Lee, Jennifer Sojka, Cody McCoy, and Venessa Alejandro. I also want to thank the many people who work at UCLA. Most especially, I would like to thank Jocelyn Yamadera for helping me in whatever way she could.

Finally, I want to thank my friends and family. My sister, Kristin, thanks for being a great big sister. My parents, Signe and Peter, have always been my biggest supporters. I wouldn't be here today without you two.

Funding for this dissertation came from the Department of Education's GAANN Fellowship, the National Science Foundation GK-12 Fellowship, and Department of Ecology and Evolutionary Biology at UCLA.

Chapter 2 is a reprint of a published paper – Petelle, M.B. & Blumstein, D.T. 2014. A critical evaluation of subjective ratings: Unacquainted observers can reliably assess certain personality traits. *Current Zoology*, 60 (2): 162-169. Chapter 3 is also a reprint of a published work – Petelle, M.B., McCoy, D.E., Alejandro, V., Martin, J.G.A., & Blumstein, D.T. 2013. Development of boldness and docility in yellow-bellied marmots. *Animal Behaviour*, 86: 1147-1154. Reprints here are with the permission of ScholarOne and Elsevier, respectively. All chapters were written with the help of others, and authors from Chapter 4 and 5 give their permission for their use here.

VITA

2006	B.S. Biology, Magna Cum Laude Arizona State University Tempe, Arizona
2006	Van Ness Fellow San Diego Zoo
2006-2009	Research assistant San Diego Zoo
2009-2011	GAANN Fellowship US Department of Education
2010-2014	Teaching Assistant Department of Ecology and Evolutionary Biology University of California, Los Angeles
2012-2013	GK-12 Teaching Fellowship National Science Foundation

PUBLICATIONS AND PRESENTATIONS

Petelle M.B. and Blumstein D.T. 2014. Development, causes, and maintenance of personality in yellow-bellied marmots. Talk presented at UCLA Ecology Evolution Pub series, Los Angeles, California, USA.

Petelle M.B., Martin J.G.A. and Blumstein D.T. 2014. Maintenance of risky personalities in yellow-bellied marmots. Talk presented at Society for Comparative and Integrative Biology, University of Texas, Austin, Texas, USA.

Petelle M.B. and Blumstein D.T. 2014. A critical evaluation of subjective ratings: unacquainted observers can reliably assess personality. *Current Zoology*, 60 (2): 162-169.

Petelle M.B., McCoy D.E., Alejandro V., Martin J.G.A., and Blumstein D.T. 2013. Development of personality: boldness and docility in yellow-bellied marmots. Talk presented at Animal Behavior Society, University of Colorado, Boulder, Colorado, USA.

Young G., Petelle M.B., Kappus S., Nonacs P., Moldwin M., and Hogue T. 2013. Increasing science communication and student engagement through the SEE-LA GK-12 program. Poster presented at Animal Behavior Society, University of Colorado, Boulder, Colorado, USA.

Petelle M.B., McCoy D.E., Alejandro V., Martin J.G.A., and Blumstein D.T. 2013 Development of boldness and docility in yellow-bellied marmots. *Animal Behaviour,* 86: 1147-1154.

Blumstein D.T., Petelle M.B., and Wey T. 2012. Defensive and social aggression: repeatable but independent. *Behavioral Ecology*, 24, 457-461.

Hensley N.M., Cook T.C., Lang M., Petelle M.B., and Blumstein D.T. 2012 Personality and habitat segregation in giant sea anemones (*Condylactis gigantean). Journal of Experimental Marine Biology and Ecology*, 426-427: 1-4.

Petelle M.B. and Blumstein D.T. 2011. Context-dependent effects of the environment on marmot personality traits. Poster presented at joint conference of the Animal Behavior Society and the International Ethological Conference, University of Indiana, Bloomington, Indiana, USA.

CHAPTER 1

GENERAL INTRODUCTION

The study of animal personality has its initial beginnings in the fields of comparative psychology and physiology (Doyle & Yule, 1959; Archer, 1973; Koolhaas *et al.*, 1999; Gosling, 2001). Comparative psychologists have long used animal temperament to understand and make connections with human personality (Gosling, 2001). Likewise, physiologists have understood that there are individual differences in stress responses that they termed coping styles (Koolhaas *et al.*, 1999). These differences in stress responses, which are seen by differences in behavior, have been described along a continuum of proactive to reactive coping styles (Koolhaas *et al.*, 1999). Despite the long history of study within these disciplines, animal personality was initially eschewed by behavioral ecologists because behavioral ecologists viewed individual variation as noise around an evolutionary optimum (Sih *et al.*, 2004).

This stance, however, changed as studies showed that this variation had consequences for both fitness and population demographics. A seminal paper by Huntingford (1976) showed that aggression in three-spined sticklebacks (*Gasterosteus aculeatus*) was correlated across contexts and that this correlation influenced fitness. In another fitting example, Armitage et al. (1986), using the same population that I studied for my dissertation research, found that yellow-bellied marmots (*Marmota flaviventris*) showed individual differences in response to mirror image stimulation tests, and these differences were connected to an individual's sociability, and hence their likelihood to remain in the natal colony and not disperse. Recruitment to natal colonies has large demographic consequences because many dispersers die and are not known to reproduce (Van Vuren & Armitage, 1994).

Although these first studies of animal personality are set in a behavioral ecological context, few studies subsequently focused on the ecological or evolutionary consequences of personality. Indeed, most studies focused on determining whether a certain taxon exhibited personality (Forkman *et al.*, 1995; Gosling, 1998; Dingemanse *et al.*, 2002; Svartberg & Forkman, 2002). This led Réale et al. (2007) to call for studies that focused on integrating both ecological and evolutionary theories of animal personality. Since Réale et al. (2007), the field has seen an influx of new research that links how personality influences evolutionary dynamics and its interaction with ecology (Boon *et al.*, 2007; Smith & Blumstein, 2008; Réale *et al.*, 2009; Cote *et al.*, 2011; Pruitt & Ferrari, 2011). The focus has now turned to understanding the proximate and ultimate causes of personality (Tinbergen, 1963; Stamps & Groothuis, 2010a), but few studies have empirically tested how personality varies across life stages, identified its causes and factors responsible for its maintenance, and studied its genetic underpinnings.

In this dissertation, I focus on a well-suited model system—yellow-bellied marmots that I studied in Colorado, USA. I examine common methodological assumptions of personality tests (Chapter 2). I investigate how personality varies across developmental stages (Chapter 3). I then study mechanisms that maintain personality variation (Chapter 4). Finally, I quantify the heritability and document the presence and magnitude of genetic correlations of personality traits (Chapter 5). Chapter 2 was published in *Current Zoology* (Petelle & Blumstein 2014), and Chapter 3 was published in *Animal Behaviour* (Petelle et al. 2013). Both of these chapters are formatted for the respective journal. They are published here with permission from those publishers. Chapter 4 and 5 were written for specific journals and are thus formatted to reflect this. All manuscripts were written with coauthors, and work was done with the help of others. These manuscripts are written to reflect their contribution.

STUDY SYSTEM AND GENERAL METHODS

Yellow-bellied marmots are large, semi-fossorial ground squirrels that are found throughout the sub-alpine regions of western North America (Frase & Hoffmann, 1980). They live in colonies that can be comprised of one or more matrilineal groups. These groups are composed of one or more adult females, their kin, and one adult male. Marmot life histories are characterized by three life stages: juveniles, or young of the year, yearlings, or those that have survived their first winter, and adults, those individuals that have survived their second winter and are reproductive (Armitage, 1991). These life stages are distinctly marked by different behaviors and strategies. Juveniles have high mortality throughout the summer and over winter, experiencing upwards of 50% mortality. Most, if not all, yearling males disperse while approximately half of the females leave their natal colony (Andersen *et al.*, 1976). Finally, adults are reproductive, but depending on the environmental factors females may not produce litters every year (Armitage & Johns, 1982) Blumstein personal communication).

Data for this dissertation were collected from a population located in the East River Valley, Gunnison, Colorado (38° 57' 29" N; 106° 59' 06" W). The population lives in and around the Rocky Mountain Biological Laboratory (RMBL), and is part of a long-term demographic and behavioral study that has been in operation since 1962 (Armitage, 1991). The valley is roughly partitioned into colonies up-valley--those north of Gothic, and colonies down-valley--those south of and including Gothic. Colonies up- and down-valley differ significantly in a number of ecological variables including date of snowmelt (Svendsen, 1974), predator presence (Monclús *et al.*, 2011), and human disturbance (Li *et al.*, 2011). As part of this study, marmots are regularly live-trapped, individually marked, and observed from a distance. For this dissertation, I specifically used trapping data collected from 2002 to 2012, open field (OF) and mirror image stimulation (MIS) tests from 2010-2012, and flight initiation distance from 2003 to 2011.

We regularly live-trapped individuals throughout the active season (mid-April through mid-September). Weather permitting, large Tomahawk traps are set at the entrance of burrows and baited with Omolene^{TM} horse food. Trapping is intensive and we trap virtually all of the individuals in the population at least once a year, and many of them many more times. We assessed one personality trait, docility, through trapping behavior. Docility is a commonly used personality trait and is defined as an individual's reaction to being trapped and handled (Réale *et al.*, 2000, 2007). Upon arrival at a trap, we dichotomously score (0/1) whether an individual alarm calls, teeth chatters, struggles in the trap, bites the cage, and hesitates to walk immediately into the trap. These scores are summed and then subtracted from the total potential score. An individual with a score of five is labeled as docile for that trapping event while an individual that receives a score of zero is non-docile or pugnacious.

After scoring docility, individuals are weighed, sexed, their ano-genital distance and left hind foot measured, and ear tags replaced if necessary (Blumstein *et al.*, 2009). We also collected blood and fecal samples for most individuals. Unique identification of each individual is imperative for studies of personality; to do this each individual is given a unique mark using a non-toxic Nyazol^{TM} fur dye. After individuals are trap-processed, they are taken to an arena, made of PVC sheeting, that measured 91.4 cm^3. A mirror was placed at the bottom of one side and covered with a sliding door. Individuals were placed inside the arena for three minutes without obstacle. This constituted the open field (OF) test. Immediately after the three minutes, the sliding door was moved to expose the mirror. The following three minutes were considered the mirror image stimulation (MIS) test. All behavior was video recorded and later quantified for the following behaviors; walk, bipedal and quadrupedal look, jump, alarm call, sniff, scratch the mirror, latency to approach the mirror, and total time at the mirror. We also quantified exploration by dividing the floor into 16 equal squares and calculating the number of squares traversed and the proportion of squares visited. All behaviors were then subjected to a Principle Component Analysis and the resulting components were retained for future analysis.

In addition to these tests, we estimated flight initiation distance (FID) intermittently from 2003 to 2012. FID is a commonly used metric of boldness (Cooper Jr, 2009; Carter *et al.*, 2012). Briefly, once we arrived at a colony, marmots were allowed to acclimate to our presence. We singled out an individual and started walking slowly (0.5 m/s) towards them while simultaneously marking when they first oriented to the observer, when they first fled, how far away they were from the burrow, and other environmental and social factors. Individuals that permitted relatively close approaches were labeled as bold, while individuals who fled at relatively longer distances, were labeled shy.

VALIDATING PERSONALITY MEASUREMENTS

Animal personality research has become one the quickest growing fields in behavioral ecology (Sih *et al.*, 2004). Indeed, it is now understood that most taxa show consistent individual differences in behavior. With this growth has come a multitude of different terminology and tests to describe these traits, but what is not often shown in these studies are whether the metrics used to measure personality are reliable and valid (Gosling, 2001; Vazire *et al.*, 2007). Validity can be measured by a number of methods, but here I measure the external validity, or how well the subjective measures correlate with the more objective behavioral codings. Furthermore, another area of emerging interest, as more projects incorporate personality into their studies, are whether certain methods can be adapted to be used by observers not familiar with the subjects in the trials. This would potentially save time because training individuals can be time intensive.

Personality is typically measured by two methods, adjective ratings and behavioral codings. Ratings are thought of as a more holistic approach and use a list of adjectives and then score the subjects on a continuum within those adjectives. Researchers who spend long periods of time observing the subjects are more likely to use adjective ratings. These include those interested in animal psychology (Buss, 1991; Gosling, 2001), primatology (Capitanio, 2008), or more generally when animals are in captivity (like zoos) (Gosling, 2001). Codings are seen as more scientifically objective because they take a mechanistic approach (Martin & Bateson, 1993) to quantifying personality. Both methods require considerable time training observers, however, ratings may be more efficient if observers can be trained on the adjectives and do not need to be familiar with the subjects they rate.

In Chapter 2, I tested whether unacquainted observers can reliably rate subjects on a predetermined list of adjectives. I then used behavioral codings to test if these ratings were externally valid. I used OF and MIS trials collected during the summer of 2010 and had two undergraduate students at UCLA, who were unfamiliar with marmots, rate those subjects on a scale using 15 adjectives. I found that over half of the adjectives were reliable; they had high repeatability. Furthermore, some of these ratings were externally valid when compared with behavioral codings. I concluded that unfamiliar raters can reliably and validly score some personality traits, most notably the active/explorative trait, and that this is partially due to how visible, or more easily distinguishable this trait is from other traits.

DEVELOPMENT OF PERSONALITY

Most personality studies are short in duration; lasting from hours to several weeks, depending on the taxa (Biro *et al.*, 2010; Stamps & Groothuis, 2010a). Therefore, there is a dearth of information about the ontogeny of personality traits. This is especially interesting because personality is most often linked to life history strategies, and these are coupled with and dependent upon life stages (Biro & Stamps, 2008; Réale *et al.*, 2009). The few studies that have investigated development of personality have done so in a laboratory setting where the environment is controlled (Bell & Stamps, 2004; Sinn *et al.*, 2008). Thus, to truly understand the causes, function, and evolution of personality we must understand how it develops in free-living animals (Stamps & Groothuis, 2010a; Stamps & Groothuis, 2010b)

In Chapter 3, I investigated the development of two personality traits, docility and boldness, and its link to different life stages in yellow-bellied marmots. These two traits have been collected consistently over the past decade and give us an ideal approach to understand development of personality. For each trait, I asked: when does an individual become repeatable? Does repeatability change across age classes? Do personality levels at one life stage predict an individual's future personality? And, is there a behavioral syndrome between the two traits and is it stable across life stages?

I found that docility and boldness differ greatly in their development. Repeatability in docility emerges during the juvenile life stage and increases into adulthood. Also, an individual's docility level as a juvenile is a strong predictor of its docility level as a yearling and adult. Boldness, conversely, is only repeatable during the yearling life stage, and boldness levels are not stable across life stages. I also found no behavioral syndrome between docility and boldness, which may be reflective of the differences in development. These findings suggest that some personality traits are coupled with life stage strategy, and that the formation of behavioral syndromes may be linked to development.

CAUSES AND MAINTENANCE OF PERSONALITY

Behavior is considered to be incredibly plastic (Sih *et al.*, 2004), but it has been shown that, in most taxa, individuals are repeatable in their behavior, and are unable to act appropriately in all situations or contexts (Gosling, 2001). Lack of plasticity therefore prevents individuals from reaching the behavioral optima in a given situation. An inability to reach situation- or context-specific behavioral optima has perplexed behavioral ecologists. This, apparently maladaptive behavior, must have an underlining cause, and multiple theoretical models have been put forth to try to explain the causes and maintenance of personality (Stamps, 2007; Wolf *et al.*, 2007; Biro & Stamps, 2008; Dingemanse & Wolf, 2010; Luttbeg & Sih, 2010; Wolf & Weissing, 2010). Three hypotheses, growth-mortality tradeoffs (Stamps, 2007; Biro & Stamps, 2008), residual reproductive success (Wolf *et al.*, 2007), and state-dependent safety (Luttbeg & Sih, 2010), are the major hypotheses tested in the literature. Support for each is equivocal.

In Chapter 4, I test all three of these hypotheses by investigating the relationship between a set of personality traits (docility and activity and exploration that was measured in two contexts) and growth (growth-mortality tradeoff), age (residual reproductive value), and body mass (state-dependent safety). Furthermore, I include date of trapping, days between trials, predator presence, time of day, sex, pedestrian presence, and trial number as other covariates to understand how these variables may influence personality between contexts.

Although there is some empirical support in the literature for all three of these hypotheses, I found no support for any in this system despite having enough power to detect them. I did, however, find that a number of environmental variables differentially influenced personality traits between contexts. The importance of environmental traits became very clear from the results of Chapter 5.

QUANTITATIVE GENETICS OF PERSONALITY

Personality traits are quite variable. This variation could be maintained by a fluctuating environment, which may lead to differential selection over time. By contrast, stabilizing or directional selection should eliminate variation. However, to understand if and how selection may maintain personality variation, one must estimate the additive genetic variance of those traits (Falconer & Mackay, 1996; Lynch & Walsh, 1998). While estimating heritability is important, trait evolution may be constrained by genetic correlations (Lande & Arnold, 1983). If present, genetic correlations can prevent personality traits from evolving to their optima as well as prevent selection from eroding variation.

In Chapter 5, I estimated the heritability, permanent environment effects, maternal effects, and year effects of docility, activity and exploration in OF and MIS tests, and sociability. I also estimated the genetic, maternal, and permanent environment correlations between among personality traits and within traits across contexts. I found small to no significant heritability in a number of personality traits as well as some small, but significant permanent environment and maternal effects. I did find a number of phenotypic correlations between traits. Underpinning these correlations were either genetic or permanent environment correlations. I found no correlations between maternal effects.

These findings suggest there is a genetic background for some of these traits, and both the shared environment (permanent environment) and maternal effects influence these traits. Genetic correlations suggest that these traits are constrained in their evolution and may be a mechanism that maintains variation. However, some phenotypic correlations are caused by the shared environment, suggesting that these traits may be adaptively coupled because of that environment. Thus, personality variation is driven and may be maintained by fluctuating selection imposed by the environment.

GENERAL CONCLUSIONS AND FUTURE DIRECTIONS

Through my dissertation research I found that personality can be linked to life stages and thus may be linked to life history strategies. I found limited effects of an individual's state on personality in this population. And, I found that some personality traits are heritable and thus have some underlying genetic basis. However, phenotypic correlations were rare and most were explained by a shared environment.

The study of animal personality is now wide spread, but there are few long-term studies of free-living animals that allow one to ask the questions presented here. Indeed, this is only the third study of heritability and genetic correlations in a wild population. Thus, more studies must be conducted to fully understand the development, causes, and heritability of traits that will help us understand the maintenance of personality variation. All too often, studies test one hypothesis at a time and focus on one Tinbergian level of analysis at a time. My dissertation research had two strengths: it asked questions at multiple levels of analysis and it evaluated multiple hypotheses (when possible).

Future work could focus on evaluating the social niche specialization hypothesis in marmots. We also need to better understand the fitness consequences of personality types.

Ultimately, such studies will allow us to understand why we find such great diversity in personality in the animal kingdom.

References

Andersen, D.C., Armitage, K.B. & Hoffmann, R.S. 1976. Socioecology of marmots: female reproductive strategies. *Ecology* **57**: 552–560.

Archer, J. 1973. Tests for emotionality in rats and mice: A review. *Anim. Behav.* **21**: 205–235.

Armitage, K.B. 1986. Individuality, social behavior, and reproductive success in yellow-bellied marmots. *Ecology* **67**: 1186–1193.

Armitage, K.B. 1991. Social and population dynamics of yellow-bellied marmots: results from long-term research. *Annu. Rev. Ecol. Syst.* **22**: 379–407.

Armitage, K.B. & Johns, D.W. 1982. Kinship, reproductive strategies and social dynamics of yellow-bellied marmots. *Behav. Ecol. Sociobiol.* **11**: 55–63.

Bell, A.M. & Stamps, J.A. 2004. Development of behavioural differences between individuals and populations of sticklebacks, *Gasterosteus aculeatus*. *Anim. Behav.* **68**: 1339–1348.

Biro, P.A., Beckmann, C. & Stamps, J.A. 2010. Small within-day increases in temperature affects boldness and alters personality in coral reef fish. *Proc. R. Soc. B Biol. Sci.* **277**: 71–77.

Biro, P.A. & Stamps, J.A. 2008. Are animal personality traits linked to life-history productivity? *Trends Ecol. Evol.* **23**: 361–368.

Blumstein, D.T., Wey, T.W. & Tang, K. 2009. A test of the social cohesion hypothesis: interactive female marmots remain at home. *Proc. R. Soc. B Biol. Sci.* **276**: 3007–3012.

Boon, A.K., Réale, D. & Boutin, S. 2007. The interaction between personality, offspring fitness and food abundance in North American red squirrels. *Ecol. Lett.* **10**: 1094–1104.

Buss, D.M. 1991. Evolutionary personality psychology. *Annu. Rev. Psychol.* **42**: 459–491.

Capitanio, J.P. 2008. Personality and disease. *Brain. Behav. Immun.* **22**: 647.

Carter, A.J., Heinsohn, R., Goldizen, A.W. & Biro, P.A. 2012. Boldness, trappability and sampling bias in wild lizards. *Anim. Behav.* **83**: 1051–1058.

Cooper Jr, W.E. 2009. Variation in escape behavior among individuals of the striped plateau lizard *Sceloporus virgatus* may reflect differences in boldness. *J. Herpetol.* **43**: 495–502.

Cote, J., Fogarty, S., Brodin, T., Weinersmith, K. & Sih, A. 2011. Personality-dependent dispersal in the invasive mosquitofish: group composition matters. *Proc. R. Soc. B Biol. Sci.* **278**: 1670–1678.

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. *Anim. Behav.* **64**: 929–938.

Dingemanse, N.J. & Wolf, M. 2010. Recent models for adaptive personality differences: a review. *Philos. Trans. R. Soc. B Biol. Sci.* **365**: 3947–3958.

Doyle, G. & Yule, E.P. 1959. Grooming activities and freezing behaviour in relation to emotionality in albino rats. *Anim. Behav.* **7**: 18–22.

Falconer, D.S. & Mackay, T.F.C. 1996. *Introduction to quantitative genetics*. (4th ed.) Longmans Green, Harlow, Essex, UK.

Forkman, B., Furuhaug, I.L. & Jensen, P. 1995. Personality, coping patterns, and aggression in piglets. *Appl. Anim. Behav. Sci.* **45**: 31–42.

Frase, B.A. & Hoffmann, R.S. 1980. Marmota flaviventris. *Mamm. Species* 1–8.

Gosling, S.D. 2001. From mice to men: what can we learn about personality from animal research? *Psychol. Bull.* **127**: 45.

Gosling, S.D. 1998. Personality dimensions in spotted hyenas (*Crocuta crocuta*). *J. Comp. Psychol.* **112**: 107.

Huntingford, F.A. 1976. The relationship between anti-predator behaviour and aggression among conspecifics in the three-spined stickleback, *Gasterosteus aculeatus*. *Anim. Behav.* **24**: 245–260.

Koolhaas, J.M., Korte, S.M., De Boer, S.F., Van Der Vegt, B.J., Van Reenen, C.G., Hopster, H., *et al.* 1999. Coping styles in animals: current status in behavior and stress-physiology. *Neurosci. Biobehav. Rev.* **23**: 925–935.

Lande, R. & Arnold, S.J. 1983. The measurement of selection on correlated characters. *Evolution* **37**: 1210.

Li, C., Monclús, R., Maul, T.L., Jiang, Z. & Blumstein, D.T. 2011. Quantifying human disturbance on antipredator behavior and flush initiation distance in yellow-bellied marmots. *Appl. Anim. Behav. Sci.* **129**: 146–152.

Luttbeg, B. & Sih, A. 2010. Risk, resources and state-dependent adaptive behavioural syndromes. *Philos. Trans. R. Soc. B Biol. Sci.* **365**: 3977–3990.

Lynch, M. & Walsh, B. 1998. *Genetics and analysis of quantitative traits*. Sinauer Associates Inc.

Martin, P. & Bateson, P.P.G. 1993. *Measuring Behaviour: An Introductory Guide*. Cambridge University Press.

Monclús, R., Tiulim, J. & Blumstein, D.T. 2011. Older mothers follow conservative strategies under predator pressure: The adaptive role of maternal glucocorticoids in yellow-bellied marmots. *Horm. Behav.* **60**: 660–665.

Petelle, M. & Blumstein, D. 2014. A critical evaluation of subjective ratings: Unacquainted observers can reliably assess certain personality trait. *Curr. Zool.* **60**: 162–169.

Petelle, M.B., McCoy, D.E., Alejandro, V., Martin, J.G. & Blumstein, D.T. 2013. Development of boldness and docility in yellow-bellied marmots. *Anim. Behav.* **86**: 1147–1154.

Pruitt, J.N. & Ferrari, M.C.O. 2011. Intraspecific trait variants determine the nature of interspecific interactions in a habitat-forming species. *Ecology* **92**: 1902–1908.

Réale, D., Gallant, B.Y., Leblanc, M. & Festa-Bianchet, M. 2000. Consistency of temperament in bighorn ewes and correlates with behaviour and life history. *Anim. Behav.* **60**: 589–597.

Réale, D., Martin, J., Coltman, D.W., Poissant, J. & Festa-Bianchet, M. 2009. Male personality, life-history strategies and reproductive success in a promiscuous mammal. *J. Evol. Biol.* **22**: 1599–1607.

Réale, D., Reader, S.M., Sol, D., McDougall, P.T. & Dingemanse, N.J. 2007. Integrating animal temperament within ecology and evolution. *Biol. Rev.* **82**: 291–318.

Sih, A., Bell, A.M., Johnson, J.C. & Ziemba, R.E. 2004. Behavioral syndromes: an integrative overview. *Q. Rev. Biol.* **79**: 241–277.

Sinn, D.L., Gosling, S.D. & Moltschaniwskyj, N.A. 2008. Development of shy/bold behaviour in squid: context-specific phenotypes associated with developmental plasticity. *Anim. Behav.* **75**: 433–442.

Smith, B.R. & Blumstein, D.T. 2008. Fitness consequences of personality: a meta-analysis. *Behav. Ecol.* **19**: 448–455.

Stamps, J.A. 2007. Growth-mortality tradeoffs and "personality traits" in animals. *Ecol. Lett.* **10**: 355–363.

Stamps, J. & Groothuis, T.G. 2010a. The development of animal personality: relevance, concepts and perspectives. *Biol. Rev.* **85**: 301–325.

Stamps, J.A. & Groothuis, T.G. 2010b. Developmental perspectives on personality: implications for ecological and evolutionary studies of individual differences. *Philos. Trans. R. Soc. B Biol. Sci.* **365**: 4029–4041.

Svartberg, K. & Forkman, B. 2002. Personality traits in the domestic dog (*Canis familiaris*). *Appl. Anim. Behav. Sci.* **79**: 133–155.

Svendsen, G.E. 1974. Behavioral and environmental factors in the spatial distribution and population dynamics of a yellow-bellied marmot population. *Ecology* **55**: 760–771.

Tinbergen, N. 1963. On aims and methods of Ethology. *Z. Für Tierpsychol.* **20**: 410–433.

Van Vuren, D. & Armitage, K.B. 1994. Survival of dispersing and philopatric yellow-bellied marmots: what is the cost of dispersal? *Oikos* **69**: 179–181.

Vazire, S., Gosling, S.D., Dickey, A.S. & Schapiro, S.J. 2007. Measuring personality in nonhuman animals. *Handb. Res. Methods Personal. Psychol.* 190–206.

Wolf, M., van Doorn, G.S., Leimar, O. & Weissing, F.J. 2007. Life-history trade-offs favour the evolution of animal personalities. *Nature* **447**: 581–584.

Wolf, M. & Weissing, F.J. 2010. An explanatory framework for adaptive personality differences. *Philos. Trans. R. Soc. B Biol. Sci.* **365**: 3959–3968.

Current Zoology 60 (2): 162–169, 2014

A critical evaluation of subjective ratings: Unacquainted observers can reliably assess certain personality traits

Matthew B. PETELLE[1*], Daniel T. BLUMSTEIN[1, 2]

[1] Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095-1606, USA
[2] Rocky Mountain Biological Laboratory, Crested Butte, CO 81224, USA

Abstract Methods to measure consistent individual differences in behavior (i.e. animal personality) fall into two categories, subjective ratings and behavioral codings. Ratings are seldom used despite being potentially more efficient than codings. One potential limitation for the use of ratings is that it is assumed that long-term observers or experts in the field are required to score individuals. This can be problematic in many cases, especially for long-term ecological studies where there is high turnover in personnel. We tested whether raters who were unacquainted with subjects could produce reliable and valid personality assessments of yellow-bellied marmots *Marmota flaviventris*. Two raters, previously unacquainted with individuals and marmot behavior, scored 130 subjects on fifteen different adjectives in both open-field (OF) and mirror image stimulation (MIS) trials. Eight OF and nine MIS adjectives were reliable as indicated by both a high degree of intra-observer and inter-observer reliability. Additionally, some ratings were externally valid, correlating with behavioral codings. Our data suggest that activity/exploration and sociability can be a reliable and valid measurement of personality traits in studies where raters were unacquainted with subjects. These traits are observable with the personality tests we used; otherwise researchers using unacquainted raters should be cautious in the tests they employ [*Current Zoology* 60 (2): 162–169, 2014].

Keywords Animal personality, Behavior codings, Marmots, Subjective ratings

Animal personality (i.e., consistent individual differences in behavior) has been documented in numerous domestic and wild species (Gosling, 2001; Réale et al., 2007) and from invertebrates to vertebrates (Gosling, 2001; Hensley et al., 2012; Mather and Logue, 2013). Behavioral codings and subjective ratings are two methods used to quantify personality (Gosling, 2001; Vazire et al., 2007). Both methods are applicable for use in the animals' home environments or in behavioral tests, such as open-field and novel object tests.

Behavioral codings measure the presence/absence, frequency, and/or duration of specific postures or behaviors, whereas subjective ratings use observers to score individuals based on a list of adjectives. In studies of non-human animals, codings are more commonly used because of their perceived objectivity and lack of human bias. Whereas ratings are used to a lesser extent, they are seen as a more holistic way to assess personality; additionally they are seen as more efficient than behavioral codings because of how quickly they can be conducted once an observer is trained (Vazire et al., 2007).

Despite the potential advantages of ratings, short- and long-term studies with high personnel turnover may not utilize this method because of the notion that raters must be well acquainted with subjects in order to accurately assess personality. Consequently, in the majority of studies that use ratings, observers are commonly breeders, trainers, or long-term animal care providers (Carter et al., 2012; Fratkin et al., 2013; Uher and Asendorpf, 2008; Wilsson and Sinn, 2012). This can be problematic for long-term ecological studies where there is high turnover in personnel. Additionally, a potential consequence of using well-acquainted observers is the potential for confirmation bias due to preconceptions that raters may have of animal subjects (Highfill et al., 2010). Surprisingly, we do not yet fully understand how acquaintance with subjects may influence ratings in either captive or wild studies.

There is research to suggest that while reliability of measures increases with level of acquaintance, raters less acquainted with subjects can also score subjects satisfactorily (Martau et al., 1985; Wemelsfelder et al., 2000). In Martau et al.'s (1985) study of 12 Japanese macaques *Macaca fuscata*, well acquainted and less acquainted raters scored individuals. Less acquainted

---

Received May 15, 2013; accepted Aug. 3, 2013.

* Corresponding author. E-mail: mpetelle@ucla.edu

raters observed subjects for up to 1 hour a day for 5 days before rating those same individuals while familiar raters observed subjects for 2 hour a day for up to a month. Although well-acquainted raters had higher inter-observer agreements, raters less familiar with the animals were still able to achieve high inter-observer agreement. However, in this case, the less acquainted raters had a level acquaintance with test subjects typically not achievable in many field studies. Wemelsfelder et al. (2000) found that multiple unacquainted observers had clear agreement in how they qualitatively described pig behavior, but these observer ratings were not tested for validity.

Before a measurement can be informative, it must be both reliable and valid. Reliability can be assessed with two methods: inter-rater agreement, and test-retest reliability (Vazire et al., 2007). Inter-rater agreement, typically measured by intra-class correlation coefficients, is an index of how well multiple observers agree in their personality ratings of an individual. Gosling (2001), in an extensive review of animal personality, found that inter-observer agreement in animals was comparable to reliability estimates in the human personality literature (grand mean 0.52). Furthermore, reliability is also assessed through test-retest reliability, or repeatability. This statistic describes how consistent an individual's personality score is across time. Repeatability depends upon taxa, sex, age, laboratory vs. field, and length between tests (Bell et al., 2009; Gosling, 2001). Gosling (2001) found that test-retest reliabilities were generally high with a range from 0.31–0.90.

Validity is an index of how well a measurement is describing what it is supposed to measure (Vazire et al., 2007). Validity can be assessed with a number of techniques. One common method to assess the external validity of ratings is to compare them to behavioral codings that are associated with that particular adjective (Gosling, 2001). For example, an individual rated as being highly sociable may spend more time at a mirror during a mirror image stimulation test or be more embedded in a social network. There are several examples of acquainted raters, up to two hours pre-trial observation, assigning subjective scores that externally predict an individuals behavioral coding in other tests (Fox and Millam, 2010; Barnard et al., 2012; Carter et al., 2012).

Here we test the reliability and validity of ratings on a long-term study of yellow-bellied marmots *Marmota flaviventris* with raters that were unacquainted with individuals and, before training, with their species-specific behavior. If subjective ratings are reliable and valid, personnel unacquainted with subjects can use them in standardized test situations.

1 Materials and Methods

1.1 Study area and system

We conducted experiments in and around the Rocky Mountain Biological Laboratory (RMBL, 38°57'N, 106°59'W), Gothic, CO, USA in 2010 (May–August). Marmots were regularly live-trapped and transferred to a cloth, handling bag where sex, reproductive status, and mass were determined. Marmots were marked with permanent ear tags for identification as well as unique fur marks (with Nyanzol fur dye) for observation from afar. Almost all marmots from the population are trapped at least once during the active season (mid-April to mid-September). Yellow-bellied marmots from this population have been previously shown to have personalities (Armitage and Van Vuren, 2003; Blumstein, et al., 2012; Svendsen and Armitage, 1973).

1.2 General tests

Open-field (OF) and mirror image stimulation (MIS) trials were conducted in an arena measuring 91.4 cm^3 made of 0.47 cm opaque PVC sheeting with a wire mesh top to prevent escape. A mirror (30.5×61.0 cm) was placed at the base of one side of the arena and covered with an opaque sliding door. A door (61.0 cm^2) was cut out of the opposite side. Sixteen (22.9 cm^2) squares were drawn on the bottom of the arena in a grid to record location of individuals. The arena was placed under a canopy for shade and to standardize the light environment. Trials were video-recorded (Sharp Mini DV Digital Camera) from above, for later scoring. We gently released individuals from the top of the arena. The first three minutes were considered the open-field test. During the open field test, individuals were allowed to freely explore the arena. The OF setup is similar to one used for testing personality in Alpine marmots *Marmota marmota* (Constantini et al. 2012; Ferrari et al., 2013; Réale personal communication). Immediately after the first three minutes, the MIS trial began by removing the sliding door to expose the mirror. Upon trial completion, we placed a Tomahawk trap within the door and urged the marmot inside. We returned individuals to the location originally trapped and cleaned the arena with a vinegar and water solution before the next trial. In total, we performed 205 open field (OF) and mirror image stimulation (MIS) trials on 130 individuals (32 juvenile females, 30 juvenile males, 20 yearling females, 22 yearling males, 16 adult females, and 10 adult males). Seventy-seven animals were tested twice, with six of those animals being tested a third time. All trials were included in analyses. Marmots were trapped opportunistically, and therefore individuals were tested sporadically throughout the active season. We used open field and mirror image stimulations because they are standardized; rating individual personality in natural settings would require raters to understand both the social and environmental context in which the behavior was recorded. Furthermore, we included juvenile individuals because they have been shown to exhibit personality (Armitage, 1986a; unpublished data).

1.3 Personality measurements

1.3.1 Subjective ratings

Videos of trials were sorted and viewed by sex and age categories to control for sex-specific ontogenetic variation in behavior. Thus, all scores were relative to the same age/sex category and not between all individuals. We chose 15 adjectives (Table 1), some from a previous list used on rhesus macaques (Capitanio, 1999), and others that have been used recently on studies of heteromyid rodents with high intraclass correlations (L. Baker, pers. comm., University of British Columbia). Marmots were scored on a scale from 1–7 in increments of 0.25, where 1 describes the individual as not exhibiting the trait, while 7 describes the trait being fully exhibited. This is similar to the method employed by Capitanio (1999), except we allowed for a finer division of ratings.

Table 1 Intra-class correlation coefficients of adjectives used to describe yellow-bellied marmots in open-field and mirror image stimulation tests

Adjective	OF		MIS	
	ICC	*P*	ICC	*P*
Active	**0.577**	**<0.0001**	**0.673**	**<0.0001**
Aggressive	-0.001	0.503	**0.511**	**<0.0001**
Apprehensive	**0.231**	**0.031**	-0.199	0.902
Cautious	0.013	0.463	-0.039	0.606
Confident	0.003	0.493	**0.271**	**0.012**
Curious	**0.577**	**<0.0001**	**0.567**	**<0.0001**
Excitable	**0.51**	**<0.0001**	**0.426**	**<0.0001**
Fearful	0.165	0.099	0.025	0.428
Irritable	0.109	0.205	0.153	0.118
Oppositional	**0.635**	**<0.0001**	**0.366**	**0.001**
Playful	0.063	0.322	**0.582**	**<0.0001**
Protective	**0.741**	**<0.0001**	**0.791**	**<0.0001**
Deliberate	0.181	0.078	0.173	0.088
Solitary	**0.697**	**<0.0001**	**0.539**	**<0.0001**
Strong	**0.264**	**0.015**	0.067	0.311

Significant values in bold.

Two raters (UCLA undergraduates) were chosen from a pool of undergraduate applicants. Neither rater had observed marmot behavior prior to watching these trials. Both raters were given the adjectives and viewed trials from juvenile, female marmots. After viewing, raters and MP discussed the adjectives and the behaviors that potentially constituted each adjective. Each rater scored 15 randomly selected juvenile female OF/MIS trials and scored them up to five times until they had high intra-rater agreement. High intra-rater agreement was defined as scores having a $r_S > 0.90$. Raters watched, but did not score, 10~15 trials of the subsequent sex/age category (e.g., juvenile females; juvenile males; yearling females, etc.) to understand differences in behavior between individuals and the previous category. All trials were watched and rated on computers at UCLA.

1.3.2 Quantitative codings

Behavior was scored using the event recorder JWatcher (Blumstein and Daniel, 2007) to calculate the number of events and the proportion of time spent walking, looking (quadrapedal and bipedal), jumping, alarm calling, smelling or sniffing, and, for MIS only, scratching, pawing, or pressing their nose against the mirror. Additionally, activity was scored by counting the number of lines crossed using the nose of the subject as an indicator of its location, proportion of squares visited, and for MIS only, the proportion of time spent in front of the mirror and on the mirrored half of the arena (Table 1). Prior to scoring trials, scorers were trained to have high intra- and inter-observer agreement ($r > 0.95$). To ensure high intra- and inter-observer agreement in quantifying behavior, MP scored a trial multiple times until the frequencies of all behaviors were equal and total durations of behaviors were within 5% between each scoring events. This method was carried out for five trials. Other scorers had to record the same behavioral frequency and estimated durations to ensure inter-observer agreement. Raters did not code behaviors. This was done by MP and other trained UCLA undergraduates.

1.4 Analyses

1.4.1 Inter-rater and test-retest reliability

All individual marmots were grouped for analysis. We analyzed OF and MIS separately. To assess inter-rater reliability for each of the 15 adjectives, we used an intraclass correlation coefficient (ICC) using a two way mixed model that measured consistency because both coders rated all individuals (Shrout and Fleis, 1979). Adjectives that had a significant ICC ($P < 0.05$) were included in future analyses. All further analyses were based on a single rating that was obtained by averaging rater scores.

We assessed test-retest reliability using individual repeatability. To obtain repeatability for individual marmots, we fit a linear mixed effects model for each adjective with age category, sex, age category * sex, rater, and trial as fixed effects, and individual as a random effect. Age category and sex have been found to influence other behaviors, including personality dimensions (Blumstein et al., 2012). We included the scores from both raters in the model and included rater as a fixed effect. We also included trial to control for habituation effects. We estimated the repeatability of each adjective by dividing the variance explained by the individual by the total phenotypic variance explained by the model. Significance of repeatability was estimated with a log-likelihood ratio test (Pinheiro and Bates, 2000). Only adjectives that had significant inter-rater and test-retest reliability were included in rating validity.

1.4.2 Validity of ratings

We tested rating external validity by including all ratings with behavioral codings in a principal component analysis. Ratings and codings that are correlated load onto the same component (J.G.A. Martin, pers. comm., University of Aberdeen). We used a Varimax rotation to aid in interpretation. For component selection, we conducted a parallel analysis with 1,000 randomly selected data sets with 95% confidence intervals for both OF and MIS PCAs. Significant components were kept for further interpretation (O'Connor, 2000). Variables with values > |0.40| were used to interpret factors. All analyses were conducted in SPSS v. 18.0 (Chicago, Il) and R 2.14.0 (R Development Core Team 2011) with the package lme4 (Bates et al., 2011). We set our alpha to 0.05.

2 Results

2.1 Inter-rater reliability and test-retest reliability

Eight of the fifteen adjectives for OF had significant ICCs. Additionally, nine of the fifteen adjectives for MIS had significant ICCs (Table 1). Six of the eight OF adjectives had significant repeatability: active (r = 0.182, LRT = 19.853, P < 0.0001), curious (r = 0.123, LRT = 7.838, P = 0.005), excitable (r = 0.170, LRT = 11.482, P = 0.0007), oppositional (r = 0.321, LRT = 40.132, P < 0.0001), protective (r = 0.369, LRT = 55.315, P < 0.0001), and solitary (r = 0.221, LRT = 23.715, P < 0.0001). All nine MIS adjectives had significant repeatability: active (r = 0.330, LRT = 38.213, P < 0.0001), aggressive (r = 0.185, LRT = 14.258; P = 0.0002), confident (r = 0.111, LRT = 5.318, P = 0.021), curious (r = 0.220, LRT = 20.872, P < 0.0001), excitable (r = 0.258, LRT = 28.894, P < 0.0001), oppositional (r = 0.156, LRT = 10.317, P = 0.001); playful (r = 0.408, LRT = 41.078, P < 0.0001), protective (r = 0.472, LRT = 75.956, P < 0.0001), and solitary (r = 0.222, LRT = 19.258, P < 0.0001).

2.2 Validity of ratings

Principle component analysis for the open field test extracted two components explaining 57.13% of the variation. The first component was interpreted as an activity and exploration factor. It was loaded with the proportion of boxes visited, number of lines crossed, number of jumps, number of rear looks, number of walks, proportion of time looking, proportion of time in rear look, proportion walking, active, curious, oppositional, protective, and solitary. The second component was also interpreted as an exploration factor with number of sniffs and proportion of time sniffing as significant variables (Table 2).

Principle component analysis for the mirror image stimulation test extracted five components explaining 69.19% of the variation. The first component was interpreted as an activity and exploration factor. It was

Table 2 Summary of principle component analysis for open-field (OF)

OF	Component	
Behaviors/adjectives	Activity/Exploration	Exploration
Active	**0.619**	0.246
Curious	**0.43**	0.285
Excitable	0.08	0.025
Oppositional	**0.575**	-0.004
Protective	**-0.782**	-0.123
Solitary	**-0.769**	-0.126
Prop boxes visited	**0.574**	0.363
N lines crossed	**0.764**	0.325
N alarm calls	-0.061	0.157
N jumps	**0.592**	-0.265
N looks	-0.117	0.387
N sniff/smell	0.254	**0.912**
N rear looks	**0.861**	-0.248
N walks	**0.795**	0.254
Prop look	**-0.905**	-0.326
Prop sniff/smell	0.135	**0.917**
Prop rear look	**0.831**	0.017
Prop walk	**0.845**	0.182

Variables with coefficients larger than |0.4| are highlighted in bold.

loaded with proportion of boxes visited, number of lines, number of looks, number of sniffs, number of walks, proportion of time looking, sniffing, and walking. Also included are active and curious adjectives. The second component was interpreted as a sociability component with proportion of time spent at the mirror, proportion of time spent on the mirrored half, number of scratches or nose touches, proportion of time scratching or nose touching, active and curious. The third component was also interpreted as a sociability component with active, aggressive, confident, oppositional, playful, protective, and solitary adjectives. The fourth component was also associated with exploration. It was loaded with number of rear looks, proportion of time looking and proportion of time rear looking. The fifth component was labeled as an excitability component with number of jumps and active, excitable, and oppositional adjectives (Table 3).

3 Discussion

Numerous studies have already found that acquainted raters can assess personality (Gosling, 2001), thus, this study investigates whether unacquainted raters can reliably and validly score personality traits. We found that subjective ratings by unacquainted raters were reliable and valid for two personality traits--activity/exploration and sociability. Specifically, subjective ratings within open-field tests were used to identify an activity/exploration personality trait while mirror image stimulation identified both an activity and a sociability personality trait. These results suggest that in certain standardized tests, subjective ratings made by people not intimately familiar with the subjects can be a useful method to quantify personality dimensions.

3.1 Reliability of personality measurements

The majority of our adjectives had significant inter-rater reliabilities. Six adjectives with significant ICCs were shared across both OF and MIS tests. This suggests that these adjectives are perhaps easier to recognize within and across situations. Active, curious, excitable, protective, and solitary were all found to have

Table 3 Summary of principle component analysis for mirror image stimulation

MIS	Component				
Behaviors/ adjectives	Activity/ Exploration	Sociability	Sociability 2	Exploration 2	Excitability
Active	0.554	0.214	0.536	0.128	0.276
Aggressive	0.125	0.1	0.52	0.121	0.689
Confident	0.221	0.201	0.756	-0.059	0.199
Curious	0.472	0.413	0.281	0.012	0.076
Excitable	0.048	0.162	-0.126	-0.05	0.753
Oppositional	0.133	-0.108	0.438	0.089	0.68
Playful	0.211	0.346	0.671	0.011	0.023
Protective	-0.375	-0.285	-0.634	-0.336	-0.127
Solitary	-0.24	-0.181	-0.604	-0.385	0.043
Prop boxes visited	0.776	0.214	0.079	0.252	0.223
N lines crossed	0.693	0.259	0.101	0.199	0.19
Prop at mirror	0.058	0.847	0.21	0.024	0.082
Prop mirror half	0.101	0.684	0.191	0.141	0.028
N alarm calls	-0.115	0.001	-0.041	-0.051	0.017
N jumps	0.075	0.18	0.078	0.381	0.652
N looks	0.48	0.246	0.103	-0.025	0.073
N sniff/smell	0.846	0.057	0.241	0.166	-0.021
N scratch/paw	0.396	0.73	0.226	0.042	0.158
N rear looks	0.298	-0.042	0.078	0.89	0.113
N walks	0.748	0.282	0.217	0.257	0.125
Prop looks	-0.516	-0.339	-0.181	-0.647	-0.118
Prop sniff/smell	0.824	0.043	0.222	0.121	-0.061
Prop scratch/smell	0.237	0.797	0.158	0.019	0.135
Prop rear look	0.234	0.019	0.063	0.883	0.118
Prop walk	0.741	0.08	0.266	0.451	0.054

Variables with coefficients larger than |0.4| are highlighted in bold.

similar, if not higher, inter-rater reliability than other studies (0.62, 0.47, 0.38, 0.38, and 0.43 respectively) (see Gosling, 2001). Other adjectives, however, may not be appropriate for all contexts, thus accounting for differences in reliability scores, or differences between individuals may be too subtle for specific tests or observers to identify (Meagher, 2009). Interestingly, only two of the five adjectives, active and curious, were observed to have high observability across species. Observability refers to how visible a trait is within a given situation or context. We are not sure why the other three adjectives were so high compared to Gosling's (2001) findings, perhaps these adjectives are appropriate for this species within this context, and are thus more observable. Open-field tests, for example, were designed to assess fear and activity, thus it is not surprising that adjectives describing these traits may be easier to rate in this situation. Conversely, both aggression and playfulness are commonly thought of as social attributes and might therefore be more observable in the mirror-image stimulation.

Of those adjectives with significant inter-rater reliabilities we found many of these to be repeatable. Personality, by definition, must be repeatable, and therefore the test-retest reliability is essential to include in any analysis of ratings. As our study shows, adjectives that have high inter-rater agreement are not necessarily repeatable, and thus should not necessarily be viewed as personality traits without further justification. Additionally, our repeatability estimates are generally moderate, but fall within the range of repeatable behaviors (Bell et al., 2009).

We should note that we did not test all individuals multiple times. While this could affect repeatability estimates for linear mixed effects models, Martin et al. (2011) advised that large data sets ($n > 200$) are sufficient to estimate individual differences, and that including individuals with one observation actually increases the power to detect these differences. Therefore, we are confident that our results accurately reflect the test-retest reliability of these adjectives.

We recognize that the use of two raters can result in an overestimate of ICC scores, and therefore our results indicate the upper-limit for reliability in these scores. However, our results suggest that just two raters can reliably score certain adjectives. Studies that use acquainted raters typically rely on one to five raters (Martau et al., 1985; Highfill et al., 2010; Barnard et al., 2012). Moreover, this experiment is part of an ongoing ecological study where high personnel turnover is common. Consequently we have a vested interest in determining if a minimum number of unacquainted raters will suffice in judging personality.

3.2 Validity of subjective ratings

Principle Component Analysis revealed that the five reliable adjectives in the OF test were correlated with behaviors that can often be used to define an activity or exploration trait. Thus, our study suggests that raters, unacquainted with subjects, were able, with minimal training, to use adjectives that describe an active/exploration personality trait during OF tests. Our results are consistent with other studies on Alpine marmots where the first component reveals an activity/exploration trait with movement and upright posture being correlated (Ferrari et al., 2013).

We also found that raters were able to describe activity/exploration within the MIS test along with a sociability component. MIS tests are widely used to assess how individuals interact with an unknown conspecific, and therefore they are often used as a metric of sociability (Armitage, 1986a; Armitage, 1986b). Additionally, we found an excitability component with aggressive, excitable, and oppositional loading significantly with number of jumps. This component was not seen in the OF test, suggesting that these correlated behaviors are related to being exposed to a mirror. Excitability has been shown in a number of studies that use ratings and is common in laboratory studies of rats (Cerbone, 1993; Gosling and John, 1998).

Interestingly, the fact that curious loads positively on two components, activity/exploration and sociability, suggests that subjective ratings provide a broader qualitative description, or holistic view of individuals, which may cover multiple traits (Uher and Asendorpf, 2008). Surprisingly, we found that adjectives that describe sociability-playful and aggressive-were not associated with time spent at the mirror. This suggests that although adjectives such as playful and aggressive can be reliably scored, they are not externally valid in this context to explore sociability.

Adjectives that were not reliably scored, or were reliable and not valid, may result from the tests not being ecologically relevant. These adjectives may be more observable (reliable and valid) if underlying tests are able to expose those underlying traits. Another potential method to pinpoint more relevant adjectives is to have them chosen to reflect traits known to exist in the test species (e.g. Armitage, 1986b and Blumstein et al., 2006). For example, mirror image stimulation codings have previously been used to determine sociability in marmots. These MIS scores were ecologically relevant, correlating with social interactions and reproductive success (Armitage and Van Vuren, 2003). Although these adjectives are useful in describing personality traits in this specific population of marmots, each species and population has different traits and correlation between those traits (Bell and Stamps, 2004; Dingemanse et al., 2007). Thus, a different set of adjectives may be a better indicator of personality traits. Taking a bottom-up approach, or watching individuals in ecologically relevant situations and then listing potential adjectives might be a more effective way of using adjectives (Uher and Asendorpf, 2008). Thus, for long-term ecological studies, personnel well acquainted with the species and individuals in the population should determine adjectives and tests used to define personality traits (Meagher, 2009). This method can potentially be used for a number of taxa including some invertebrates given that the personality traits are highly observable in a standardized test. For example, it may be very easy for unacquainted observers to rate individuals on an activity/exploration axis in an open field test.

Our study suggests that projects with high personnel turnover should be able to effectively use ratings to reduce time and resources to score behaviors and quantify some personality traits provided that raters are properly trained beforehand and subjects are tested in a standardized manner. Those traits studied, however, should be restricted to ones that are explicitly observable. For example, our study shows that OF and MIS tests can be used to identify active and active/sociable traits, but not other traits. Indeed, the reliability of difficult to score traits should be generally scrutinized when relying on expert raters.

Acknowledgments M.B.P. was supported by a GAANN fellowship, a GK-12 fellowship, and the UCLA Department of Ecology and Evolutionary Biology. D.T.B. is supported by the NSF. We thank Julien Martin for statistical help, Liv Baker for additional assistance, and Lynn Fairbanks and Peter Nonacs for comments on previous drafts. We also thank Josh Ryan and Laurel Brewer for rating the individuals, and all the marmoteers, especially Leon Chong, who helped collect data. We also thank three anonymous reviewers for their comments that were helpful in the revision of this manuscript.

References

Armitage KB, 1986a. Individual differences in the behavior of juvenile yellow-bellied marmots. Behav. Ecol. Sociobiol. 18: 419–424.

Armitage KB, 1986b. Individuality, social behavior, and reproductive success in yellow-bellied marmots. Ecology 67: 1186–1193.

Armitage KB, Van Vuren DH, 2003. Individual differences and reproductive success in yellow-bellied marmots. Ethol. Ecol. Evol. 15: 207–233.

Barnard S, Siracusa C, Reisner I, Valsecchi P, Serpell JA, 2012. Validity of model devices used to assess canine temperament in behavioral tests. App. Anim. Behav. Sci. 138: 79–87.

Bates D, Maechler M, Bolker B, 2011. lme4: Linear mixed–effects models using S4 classes. R package version 0.999375-42. http://CRAN.R-project.org/package=lme4

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

Bell AM, Stamps JA, 2004. Development of behavioural differences between individuals and populations of sticklebacks *Gasterosteus aculeatus*. Anim. Behav. 68: 1339–1348.

Blumstein DT, Daniel JC, 2007. Quantifying behavior the JWatcher way. Sunderland: Sinauer Associates Inc.

Blumstein DT, Holland BD, Daniel JC, 2006. Predator discrimination and 'personality' in captive Vancouver Island marmots *Marmota vancouverensis*. Anim. Conserv. 9: 274–282.

Blumstein DT, Petelle MP, Wey TW, 2012. Defensive and social aggression: Repeatable but independent. Behav. Ecol. 24: 457–461.

Capitanio JP, 1999. Personality dimensions in adult male rhesus macaques: Prediction of behaviors across time and situation. Am. J. Primatol. 47: 299–320.

Carter A, Marshall H, Heinsohn R, Cowlishaw G, 2012. Evaluating animal personalities: Do observer assessments and experimental tests measure the same thing? Behav. Ecol. Sociobiol. 66: 153–160.

Cerbone A, Pellicano P, Sadile AG, 1993. Evidence for and against the naples high- and low-excitability rats as genetic model to study hippocampal functions. Neurosci. Biobehav. Rev. 17: 295–303.

Constantini D, Ferrari C, Pasquaretta C, Cavallone E, Carere C et al, 2012. Interplay between plasma oxidative status, cortisol and coping styles in wild alpine marmots *Marmota marmota*. J. Exper. Biol. 215: 374–383.

Dingemanse NJ, Wright J, Kazem AJ, Thomas DK, Hickling R et al., 2007. Behavioural syndromes differ predictably between 12 populations of three-spined stickleback. J. Anim. Ecol. 76: 1128–1138.

Ferrari C, Pasquaretta C, Carere C, Cavallone E, von Hardenberg A et al., 2013. Testing for the presence of coping styles in a wild mammal. Anim. Behav. 85: 1385–1396.

Fox RA, Millam JR, 2010. The use of ratings and direct behavioural observation to measure temperatment traits in cockatiels *Nymphicus hollandicus*. Ethology 116: 59–71.

Fratkin JL, Sinn DL, Patall EA, Gosling SD, 2013. Personality consistency in dogs: A meta-analysis. PLOS One 8(1): doi: 10.1371/journal.pone.0054907

Gosling SD, John OP, 1998. Personality dimensions in nonhuman animals: A cross-species review. Curr. Dir. Psychol. Sci. 8: 69–75.

Gosling SD, 2001. From mice to men: What can we learn about personality from animal research? Psychol. Bull. 127: 45–86.

Highfill F, Hanbury D, Kristiansen R, Kuczaj S, Watson S, 2010.

Rating *vs*. coding in animal personality research. Zoo Biol. 29: 509–516.

Hensley NM, Cook TC, Lang M, Petelle MB, Blumstein DT, 2012. Personality and habitat segregation in giant sea anemones *Condylactis gigantea*. J. Exp. Marine Biol. Ecol. 1–4: 426–427.

Mather JA, Logue D, 2013. The bold and the spineless: Invertebrate personalities. In: Carere C, Maestripieri D ed. Animal Personalities: Behavior, Physiology, and Evolution. Chicago: The University of Chicago Press, 23–53.

Martau PA, Caine NG, Candland DK, 1985. Reliability of the emotions profile index, primate form, with *Papio hamadryas*, *Macaca fuscata*, and two *Saimiri* species. Primates 26: 501–505.

Martin JGA, Nussey DH, Wilson AJ, Réale D, 2011. Measuring individual differences in reaction norms in field and experimental studies: A power analysis of random regression models. Methods Ecol. Evol. 2: 362–374.

Meagher RK, 2009. Observer ratings: Validity and value as a tool for animal welfare research. Appl. Anim. Behav. Sci. 119: 1–14.

Murphy LB, 1978. The practical problems of recognizing and measuring fear and exploration behaviour in the domestic fowl. Anim. Behav. 26 (Part 2): 422–431.

O'Connor BP, 2000. SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. Behav. Res. Meth. Instrum. and Comp. 32: 396–402.

Pinheiro JC, Bates DM, 2000. Mixed-effects Models in S and S-PLUS. New York: Springer Verlag.

Réale D, Reader SM, Sol D, McDougall PT, Dingemanse NJ, 2007. Integrating animal temperament within ecology and evolution. Biol. Rev. 82: 291–318.

Shrout PE, Fleis JL, 1979. Intraclass correlations: Uses in assessing rater reliability. Psychol. Bull. 86: 420–428.

Svendsen GE, Armitage KB, 1973. Mirror-image stimulation applied to field behavioral studies. Ecology 54: 623–627.

Uher J, Asendorpf JB, 2008. Personality assessment in the great apes: Comparing ecologically valid behavior measures, behavior ratings, and adjective ratings. J. Res. Pers. 42: 821–838.

Vazire S, Gosling SD, Dickey AS, Schapiro SJ, 2007. Measuring personality in nonhuman animals. In: Robins RW, Fraley CR, Krueger RF ed. Handbook of Research Methods in Personality Psychology. New York: The Guilford Press, 190–206.

Wemelsfelder F, Hunter EA, Mendl MT, Lawrence AB, 2000. The spontaneous qualitative assessment of behavioural expressions in pigs: First explorations of a novel methodology for integrative animal welfare measurement. Appl. Anim. Behav. Sci. 67: 193–215.

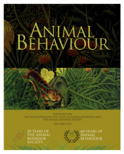
Wilsson E, Sinn DL, 2012. Are there differences between behavioral measurement methods? A comparison of the predictive validity of two ratings methods in a working dog program. Appl. Anim. Behav. Sci. 141: 158–172.

Animal Behaviour 86 (2013) 1147–1154

Contents lists available at ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav



Development of boldness and docility in yellow-bellied marmots

Matthew B. Petelle [a,*], Dakota E. McCoy [b,c], Vanessa Alejandro [c], Julien G. A. Martin [a,d], Daniel T. Blumstein [a,c]

[a] *Department of Ecology & Evolutionary Biology, University of California Los Angeles, CA, U.S.A.*
[b] *Department of Ecology & Evolutionary Biology, Yale University, New Haven, CT, U.S.A.*
[c] *The Rocky Mountain Biological Laboratory, Crested Butte, CO, U.S.A.*
[d] *School of Biological Sciences, University of Aberdeen, Aberdeen, U.K.*

ARTICLE INFO

Article history:
Received 29 January 2013
Initial acceptance 15 March 2013
Final acceptance 3 September 2013
Available online 18 October 2013
MS. number: A13-00087R

Keywords:
behavioural syndrome
boldness
docility
Marmota flaviventris
ontogeny
personality
yellow-bellied marmot

Personality traits are important because they can affect individual survival as well as how a population may respond to environmental change. How these traits arise, whether they are maintained throughout ontogeny, and how environmental factors differentially affect them throughout life is poorly understood. Understanding these pathways is important for determining the function and evolution of animal personality. We examined the development of two commonly studied personality traits, boldness and docility, in a long-term study of yellow-bellied marmots, *Marmota flaviventris*. Using data collected between 2002 and 2011, we quantified the repeatability within three age groups (juveniles, yearlings and adults), the correlation between age classes, and the behavioural syndromes of these two traits within the three life stages. We quantified boldness through flight initiation distance (FID) tests, and we quantified docility through marmots' response to being trapped. We found that boldness was repeatable only in yearlings, but docility was repeatable in all age classes. We also found that juvenile docility predicted later docility. We also found no behavioural syndrome between boldness and docility in any life stage. This suggests an adaptive hypothesis: that these personality traits develop independently and at potentially age-appropriate times. Thus, the development of personality traits may facilitate animal's coping with age-dependent requirements and constraints.

Personality, which is inferred from consistent individual differences in behaviour, can have a profound effect on a population's evolutionary and ecological dynamics (Dingemanse et al. 2004; Réale et al. 2007; Pruitt et al. 2012). Most studies of animal personality, however, focus on short time periods or one life stage. This focus provides a limited view of how personality interacts with physiology, life stage, experience and the environment (Stamps & Groothuis 2010). Indeed, an ontogenetic perspective provides a more comprehensive understanding of the function and evolution of personality (Stamps & Groothuis 2010).

Previous studies on the development of personality have focused primarily on quantifying repeatability with maternal effects (Groothuis et al. 2008; Rödel & Meyer 2011), early experience (DiRienzo et al. 2012), or across life stages (Bell & Stamps 2004; Sinn et al. 2008; Gyuris et al. 2012; Wilson & Krause 2012). For example, threespine sticklebacks, *Gasterosteus aculeatus*, were tested for stability in multiple personality traits across life stages (Bell & Stamps 2004). The authors found little stability in single personality traits across life stages, but that a syndrome, boldness and aggression, was consistently detected across ontogeny (Bell & Stamps 2004). In that study, stability was inferred from a positive correlation between juvenile and adult personality traits. Thus, individuals with stable personality traits maintained the same level of aggression or boldness relative to others across ontogeny. Studies that focused on repeatability within life stages showed mixed results. Individual dumpling squid, *Euprymna tasmanica*, were repeatable in their level of boldness as juveniles, but this repeatability disappeared for a period upon sexual maturation and reappeared in adulthood (Sinn et al. 2008). During the period of sexual maturity, and depending on their behavioural type, squid were more plastic in certain ecologically relevant contexts. For example, shyer individuals were more plastic in a feeding context than bolder individuals. Conversely, repeatability was present within life stages in both firebugs, *Pyrrhocoris apterus* (Gyuris et al. 2012) and lake frogs, *Rana ridibunda* (Wilson & Krause 2012).

These studies suggest that personality traits, in some species, are plastic within or between life stages. Thus, juvenile personality may not always predict adult personality, and plasticity may

---

[*] Correspondence: M. B. Petelle, Department of Ecology & Evolutionary Biology, University of California, 621 Charles E. Young Drive South, Los Angeles, CA 90095-1606, U.S.A.
E-mail address: mpetelle@ucla.edu (M. B. Petelle).

0003-3472/$38.00
http://dx.doi.org/10.1016/j.anbehav.2013.09.016

increase within a life stage whereby individuals may change their level of a particular personality trait. This low repeatability might be seen in certain life stages because individuals may experience differential selection over time, selection may favour plasticity (opportunity for multiple alternative strategies), or there may be a single best strategy for a given a set of environmental conditions (Fox & Westneat 2010). Thus, it is important to understand the development of personality and how these traits interact within and between life stages.

Here we capitalize on a long-term ecological study of yellow-bellied marmots, *Marmota flaviventris*, to investigate repeatability and the stability (i.e. correlation at the individual level) within and across three life stages for boldness and docility traits. Previous research on yellow-bellied marmots has shown that personality traits exist in both yearling and adult individuals and can influence fitness (Svendson & Armitage 1973; Svendson 1974; Armitage 1986; Armitage & Van Vuren 2003). Thus, we hypothesize that all three age categories will exhibit these two personality traits. We do not have a priori hypotheses about the stability of these traits across ontogeny or whether there is a behavioural syndrome between the two. Additionally, our new and extensive data set allows us to understand how environmental variables affect personality at different life stages. Thus, our first aim was to understand how repeatability varies between life stages. Our second aim was to investigate whether juvenile personality levels predict yearling and adult personality levels (that is, the longitudinal stability of these personality traits). Our third aim was to describe whether a behavioural syndrome exists and is stable between boldness and docility across these same life stages. Our last aim was to explore how an individual's current state and environment affect personality within life stages (Brydges et al. 2008; Luttbeg & Sih 2010).

METHODS

Study Subjects and Site

Between 2002 and 2011, we studied yellow-bellied marmots in the Upper East River Valley, in and around the Rocky Mountain Biological Laboratory (RMBL) in Gothic, Colorado, U.S.A. (38°77′N, 106°59′W). Yellow-bellied marmots are facultatively social, sciurid rodents, which weigh 2–6 kg, live in matrilineal groups and dwell in subalpine meadows, slopes and clearings (Frase & Hoffmann 1980). The Upper East River valley is divided in two parts, up- and down-valley, that differ in elevation, phenology and human disturbance. Three of the colonies were located up-valley while four were down-valley. Marmots were trapped on a regular basis using Tomahawk live traps and individually marked with numbered eartags for permanent identification and fur dye to facilitate identification from afar (Armitage 1982; Blumstein et al. 2009). Almost all of the individuals were trapped for the first time as juveniles and thus were of known age. Juveniles are individuals in their first summer of life; yearlings, or 1-year-olds, are in their second summer of life; adults are 2 years or older (Armitage & Downhower 1974).

Quantifying Environmental Factors

Marmots are seasonally active and we study them in a valley that is used for summer tourism. We quantified human presence for 12 days during the peak summer months of 2010 by continuously recording pedestrian activity within 300 m of six colonies (Li et al. 2011). Human traffic was not quantified in some colonies where personality data were collected, but our previous study showed that more humans were present at our down-valley sites (including those sites where impact was not specifically quantified) than our up-valley sites. Thus, we averaged human visits in two down-valley colonies and four up-valley colonies, and used these average values for unquantified colonies (down valley: Bench, Avalanche and River); we believe that these averages provide a good approximation of human disturbance in unstudied colonies. We used these averages for all years because, while not precisely quantified, human disturbance seemed not to change much between years (hiking and biking trails remained the same throughout the duration of data collection for this study; tourism was steady).

Predator presence was quantified from 2002 to 2011 by dividing the number of predators seen during observations at a colony by the total number of observation sessions at that colony. We used predator sightings only during the early season (mid-April through June) because predators become harder to view as vegetation grows during the active season. We also quantified the number of predators seen per hour of observation to test whether there were any differences between indexes. The indexes were highly correlated (Pearson correlation: $r_{95} = 0.961$, $P < 0.001$), indicating no bias in the number of predators seen per observation session. A total of 203 aerial predators and 292 terrestrial predators were seen during this time.

Quantifying Boldness with Flight Initiation Distance

To assess boldness, we conducted 563 flight initiation distance (FID) experiments on 237 individuals (86 juveniles, 81 yearlings, 70 adults) from 2003 to 2011. FID is the distance at which an individual first flees from an approaching human (Ydenberg & Dill 1986; Blumstein 2003) and is an antipredator behaviour that is commonly used as a metric for individual boldness (Cooper 2009). It should be noted that FID is inversely related to boldness: bold individuals have short FIDs whereas shy ones have large FIDs. Thus, to obtain an index that was positively related to boldness, we used the opposite of the FID (i.e. we made the values negative).

After arriving at a site, the researcher sat and quietly observed and identified subjects for at least 5 min. Once a target subject was identified, the observer walked directly towards the marmot at a constant pace. Observers were trained until they consistently walked at a 0.5 m/s pace across a variety of terrains (Runyan & Blumstein 2004). We recorded when an individual first raised its head and looked towards the researcher (alert distance), when it first fled (FID), its distance from the researcher when the trial began (start distance), its distance from a burrow when it fled (burrow distance) and its initial behaviour (forage, look, other; 'look' implies that the marmot was looking at the researcher). All distances were first marked with flags dropped during the trial and calculated afterwards by pacing. Individual pace length was calculated during training. The researcher waited at least 10 min before conducting another trial on a different subject. Individual marmots were tested no more than once per observation session (morning/afternoon sessions).

Quantifying Docility with Trapping Behaviour

Docility was assessed during 8217 trapping events for 861 juveniles, 445 yearlings and 266 adults from 2002 to 2011. We use the description of docility set forth by Réale et al. (2000). Docility here is defined as an individual's reaction to being trapped and handled. Docility is a commonly used metric of personality and is often used as a measure of risky behaviour (Réale et al. 2007, 2009; Careau et al. 2010). When trapped, marmots were transferred to a cloth handling bag for subsequent processing. At each trapping event, we recorded each marmot's behaviour while in the trap prior to being put in the trap bag. We dichotomously (i.e. 0/1) scored whether or not individuals emitted alarm calls, tooth chattered, struggled in the trap, bit the cage, and whether they failed to walk immediately into the handling bag. Following Réale et al. (2000), we summed the dichotomously scored behaviours and subtracted this from the total potential score. A score of 0 thus indicates a nondocile individual, and inversely a score of 5 indicates a docile individual.

Statistical Analyses

Age-specific repeatability and environmental effects

We first analysed each life stage separately to estimate age-specific repeatability of the behaviours and to determine the environmental effects specific at each age class. We fitted univariate linear mixed-effects models (i.e. one dependent variable with multiple fixed and random effects; Dingemanse & Dochtermann 2013) for both behaviours (boldness and docility) for each age class (juvenile, yearling, adult). For all models, individual identity and year were fitted as random effects to assess both personality and yearly environmental variation, respectively. Repeatability was estimated as the ratio of the variance associated with the individual identity effect divided by the total phenotypic variance (i.e. sum of individual, yearly and residual variances), a significant repeatability indicating personality. Additional fixed effects were fitted and are described below.

Fixed effects of flight initiation distance (boldness) included trial number per individual, trials done each day at the colony level, time (days) between trials, start and alert distances, the distance from the marmot to the burrow, sex, time of day (morning or afternoon), pedestrian traffic, predator presence, estimated mass on 15 August, estimated mass gain from 1 June to 15 August, initial behaviour and date. Trial number and the number of trials conducted at a colony per day were included to control for potential habituation effects. Time between trials was included to control for potential biases in repeatability (Bell et al. 2009). We included start distance and alert distance and the initial distance to a burrow because previous research has shown that they affect FID (start and alert: Blumstein 2010; distance to refuge: Dill & Houtman 1989). We included mass at 15 August and mass gain as proxies for body condition and growth rate, respectively. Both measures have theoretically been shown to influence personality (Stamps 2007; Biro & Stamps 2008; Luttbeg & Sih 2010). We used mass at capture as a measure of body condition. We included initial behaviour because it could influence when individuals become alert to possible predators. To control for effects from the progression of the day and season, we included time of day and the date. Flight initiation distance was square-root transformed prior to analyses to conform to the normality of residuals assumption of linear models. Fixed effects of docility were time (in days) between trapping, sex, time of day (morning or afternoon), date, pedestrian traffic, predator presence, mass gain and mass at capture. Sex was included to control for potential differences between males and females. Time of day and date were included because of differences within day and throughout the active season.

We used a log-likelihood ratio test (LRT, estimated as minus twice the difference in the likelihood of the nested models) to determine the significance of random effects between models with and without a given random effect (Pinheiro & Bates 2000). The LRT statistic follows a chi-square distribution with the difference in the number of parameters between the two models as the degree of freedom (Pinheiro & Bates 2000). Univariate linear mixed-effects models were fitted in R 2.14 (R Development Core Team 2011) with the lmer function in the lme4 package (Bates et al. 2012). We report full models after extracting parameter estimates and MCMC P values using the pvals.fnc function in the package language R (Baayen et al. 2008) based on a Markov-chain Monte Carlo sampling with 30 000 simulations.

Correlation between life stages and between behaviours

To estimate the correlations, or stability, between juvenile, yearling and adult behaviours at the individual level, we fitted trivariate mixed models (i.e. three dependent variables with multiple fixed and random effects; Dingemanse & Dochtermann 2013) for each behaviour, considering each age class as a different trait. Individual identity and year were fitted as random effects with an unstructured 3 × 3 (co)variance matrix estimating three variance components (one for each trait) and their three pairwise covariances. Correlations were then calculated from the (co)variance matrix as the rescale covariances (i.e. covariance between two traits divided by the square root of the product of the variances of the two traits). The residual matrix was constrained to be a 3 × 3 diagonal matrix because of the structure of our data. To test whether variance components differed between the three age classes, we used an LRT between models with and without constraints of equality of variance components for the three age classes. We used a similar test for pairwise comparison of variance components but constraining only two age classes at a time to be equivalent. Significance of behavioural correlations at the individual level between age classes was estimated using an LRT between models with and without a covariance parameter constrained to zero.

To identify the presence of age-specific behavioural syndromes, we estimated the correlation between both behaviours at the individual level using bivariate mixed models of boldness and docility for each age category. Individual identity and year were fitted as random effects. Each random effect was specified with an unstructured 2 × 2 (co)variance matrix thus estimating two variances (one for each behaviour) and their covariance. The residual variance matrix was fitted as a diagonal matrix (i.e. covariance fixed to zero) because of the structure of our data. The significance of the behavioural syndrome was estimated using an LRT between models with and without the covariance between docility and boldness fixed to zero.

Only fixed effects that were significant in univariate analyses were included in multivariate analyses to avoid overparametrization of models and facilitate convergence. The sample size for multivariate models is a combination of the sample size reported in the Table 1 for univariate analysis (i.e. sample size for trivariate analysis of boldness is the sum of the three sample sizes for age-specific models of boldness; see Results). Correlations estimated in multivariate analysis, however, were estimated at the individual level, meaning that only individuals with data for both traits provided information for the correlation. We thus report the number of individuals as the sample size for correlation estimates. Multivariate analyses were fitted using ASREML-R 3.0 (Gilmour et al. 2009).

Ethical Note

Marmots were studied under protocols approved by the Animal Use and Care Committees of the University of California Los Angeles and the RMBL (UCLA Protocol No. 2001-191-01 renewed annually), and under permits from the Colorado Division of Wildlife (TR917 issued annually). After trapping, individuals were released immediately at the trap location. Marmots were in traps no longer than 2–3 h, and typically much less time. Traps were shaded with vegetation on warm days. Marmot handling was brief (typically 5–15 min depending upon what data needed to be collected) and marmots were not injured during this handling. All marmots were handled while inside of a cone, cloth handling bag to reduce stress. We swabbed ears with alcohol before tagging individuals to reduce the chance of infection. FID trials are a widely used measure of risk assessment that causes only a transient change in behaviour.

Table 1
Fixed effects explaining variation in the univariate model of boldness (quantified as the negative of flight initiation distance) for three age classes of yellow-bellied marmots

	Juveniles		Yearlings		Adults	
	Estimate (SE)	*P*	Estimate (SE)	*P*	Estimate (SE)	*P*
Intercept	−1.192 (2.544)	0.238	**−7.328 (1.628)**	**<0.001**	**−6.237 (1.394)**	**<0.001**
Time between trials (days)	−0.003 (0.012)	0.595	−0.022 (0.012)	0.168	0.017 (0.011)	0.151
Trial at colony (per day)	−0.035 (0.041)	0.718	−0.060 (0.045)	0.322	0.054 (0.064)	0.388
Sex (male)	0.201 (0.208)	0.442	**0.200 (0.275)**	**0.047**	**−0.849 (0.336)**	**0.01**
Trial per individual	0.032 (0.132)	0.853	**0.120 (0.052)**	**0.005**	**0.255 (0.058)**	**<0.001**
Time (afternoon)	**0.442 (0.227)**	**0.035**	−0.089 (0.169)	0.626	0.017 (0.190)	0.915
Pedestrian	0.021 (0.017)	0.118	**0.048 (0.019)**	**0.001**	**0.031 (0.011)**	**0.003**
Mass gain	0.002 (0.001)	0.367	**0.001 (0.001)**	**0.022**	4.332e-04 (0.000)	0.16
Mass in August	−0.002 (0.001)	0.295	**−0.001 (0.000)**	**0.01**	2.62e-04 (0.000)	0.224
Predator presence	**3.614 (1.562)**	**0.005**	**6.917 (1.328)**	**<0.001**	**2.824 (0.976)**	**<0.002**
Date	−0.002 (0.010)	0.936	0.015 (0.007)	0.065	−0.002 (0.007)	0.683
Alert distance	**−0.061 (0.006)**	**<0.001**	**−0.043 (0.003)**	**<0.001**	**−0.035 (0.002)**	**<0.001**
Distance to burrow	−0.001 (0.022)	0.655	−0.014 (0.008)	0.212	**−0.044 (0.012)**	**<0.001**
Initial behaviour (look)	0.302 (0.328)	0.704	0.041 (0.194)	0.881	−0.048 (0.230)	0.772

Significant effects are in bold.

RESULTS

Environmental Factors

Environmental factors affected boldness of juveniles, yearlings and adults differently. For juveniles, boldness increased with predator pressure and as the day progressed. As expected, juveniles fled sooner if they alerted sooner (Table 1). Yearling boldness increased with predator pressure. Yearlings also became bolder as the number of trials and pedestrian traffic increased. Males were bolder than females, and individuals that gained mass quicker were also bolder. Individuals in better body condition were less bold. Again, individuals fled sooner if they alerted to the observer sooner. Adult boldness increased as trial number, pedestrian traffic and predator presence increased. Adult males were less bold than females. Furthermore, both alert distance and distance to burrow affected boldness: individuals that alerted sooner and that were farther from a burrow were less bold (Table 1).

We found that juveniles became more docile as the active season progressed (Table 2), while no such effect was seen in yearlings or adults (Table 2). Additionally, faster-growing yearlings were more docile, and yearlings in better body condition were less docile in the afternoon. Adults were more docile as the time between trials (days) increased. Males were less docile than females, and individuals were less docile in the afternoon and in areas with higher pedestrian traffic. Date and mass at capture were highly correlated for both juveniles and adults (Pearson correlation: juvenile: $r_{4006} = -0.884$; yearling: $r_{2292} = -0.894$), which created a multicollinearity issue, and therefore reduced our ability to isolate their independent effects.

Repeatability within and Correlation between Life Stages

The repeatability estimates for boldness differed significantly between the three age classes ($\chi^2_2 = 6.01, P = 0.049$). Juvenile and adult marmots were not differentially consistent in their boldness (repeatability: $r = 0.037$ and $r = 0.048$, respectively), but yearlings ($r = 0.402$) showed consistent individual differences in boldness (Table 3). The individual variance component for yearlings did not differ significantly from that for juveniles ($\chi^2_1 = 2.237, P = 0.134$), but was significantly higher than that for adults ($\chi^2_1 = 4.431, P = 0.035$), and juvenile and adult variance components did not differ significantly from one another ($\chi^2_1 = 0.001, P = 0.966$; Table 4). We found no significant correlations between age classes in boldness (juvenile–yearling: $\chi^2_1 = 0.12$, $P = 0.728$, $N = 24$ individuals; juvenile–adult: $\chi^2_1 = 0.12, P = 0.727, N = 2$ individuals; yearling–adult: $\chi^2_1 = 0.055$, $P = 0.814$, $N = 13$ individuals; Table 4, Fig. 1). The few individuals measured at different ages, however, limited our ability to correctly estimate these correlations.

Docility was repeatable in juveniles ($r = 0.168$), yearlings ($r = 0.262$) and adults ($r = 0.272$) (Table 3). The individual variance component estimates for docility differed significantly between the three age classes ($\chi^2_2 = 8.57$, $P = 0.013$). The individual variance component for adults was significantly higher than that for juveniles ($\chi^2_1 = 8.48$, $P = 0.003$), whereas the yearling component did not differ statistically from that of juveniles ($\chi^2_1 = 2.11$, $P = 0.137$) or adults ($\chi^2_1 = 2.43$, $P = 0.113$) (Table 4). Docility was significantly correlated across all life stages (juvenile–yearling: $\chi^2_1 = 64.49$, $N = 401$ individuals, $P < 0.001$; juvenile–adult: $\chi^2_1 = 13.04, N = 146$ individuals, $P = 0.001$; yearling–adult: $\chi^2_1 = 37.81$, $N = 158$ individuals, $P < 0.001$) with correlation coefficients higher than 0.6 (Table 4, Fig. 1).

Table 2
Fixed effects explaining variation in the univariate model of trapping behaviours, a measure of docility, for three age classes of yellow-bellied marmots

	Juveniles		Yearlings		Adults	
	Estimate (SE)	*P*	Estimate (SE)	*P*	Estimate (SE)	*P*
Intercept	**1.206 (0.399)**	**<0.001**	**3.851 (0.271)**	**<0.001**	**4.205 (0.155)**	**<0.001**
Time between trials (days)	−0.002 (0.002)	0.206	−1.42e-04 (0.002)	0.950	**0.004 (0.001)**	**0.005**
Sex (male)	0.022 (0.043)	0.352	−0.010 (0.063)	0.409	**−0.238 (0.086)**	**0.001**
Time (afternoon)	0.043 (0.035)	0.342	**−0.085 (0.035)**	**0.014**	**−0.137 (0.036)**	**<0.001**
Date	**0.015 (0.002)**	**<0.001**	0.003 (0.002)	0.179	1.76e-04 (0.001)	0.973
Mass gain	−1.57e-05 (3.74e-05)	0.820	**2.45e-04 (1.52e-04)**	**0.037**	4.30e-05 (3.72e-05)	0.176
Mass at capture	−1.35e-04 (8.60e-05)	0.282	**−1.43e04 (6.18e-05)**	**0.042**	3.82e-05 (4.05e-05)	0.301
Pedestrians	0.003 (0.002)	0.213	0.005 (0.004)	0.165	**−0.007 (0.004)**	**0.009**
Predator presence	4.00e-04 (0.001)	0.822	−5.03e-04 (0.001)	0.680	0.001 (0.001)	0.566

Significant effects are in bold.

Table 3
Variance, ratio and significance of random effects using log-likelihood ratio tests (LRT) for both docility and boldness univariate calculated from models for juvenile, yearling and adult yellow-bellied marmots

Trait	Number of observations, individuals	Mean (SD) traits	Variance			Ratio		LRT (*df*=1)		*P*	
			Identity	Year	Phenotypic	Identity	Year	Identity	Year	Identity	Year
Docility											
Juveniles	3316, 861	4.183 (0.967)	0.151	0.029	0.895	0.168	0.032	119.498	37.212	<0.001	<0.001
Yearlings	2294, 445	4.384 (0.870)	0.208	0.027	0.792	0.262	0.034	185.344	10.368	<0.001	0.001
Adults	2607, 266	4.294 (0.989)	0.254	0.005	0.932	0.272	0.005	421.061	5.15	<0.001	0.023
Boldness											
Juveniles	126, 86	−27.854 (24.396)	0.261	6.164	7.075	0.037	0.871	1.392	17.416	0.238	<0.001
Yearlings	204, 81	−43.268 (32.793)	0.617	0.082	1.536	0.402	0.053	13.907	0.691	<0.001	0.406
Adults	233, 70	−46.999 (38.654)	0.083	0.114	1.703	0.048	0.067	0.941	4.246	0.332	0.039

Note that repeatability estimates from trivariate (Table 4) and univariate (Table 3) models differed slightly because only significant fixed effects from the univariate models were included in the trivariate analyses.

Behavioural Syndrome

Using a bivariate model for each age class, we found that the correlation between boldness and docility was not significant for any age class (repeatability ± SE: juveniles: $r = 0.582 \pm 0.352$, LRT = 2.28, $N = 96$ individuals, $P = 0.131$; yearlings: $r = 0.186 \pm 0.193$, $\chi^2_2 = 0.829$, $N = 99$ individuals, $P = 0.362$; adults: $r = 0.587 \pm 0.936$, $\chi^2_2 = 1.608$, $N = 80$ individuals, $P = 0.204$).

DISCUSSION

Our examination of the ontogeny of two personality traits, boldness and docility, in yellow-bellied marmots found four main results that have general implications for our understanding of the development of personality. First, environmental effects differed for each age class. Second, personality emerged in different age classes. Third, personality in one age class was not necessarily correlated with personality in another. Fourth, a behavioural syndrome was not present in any age class.

Environmental Effects

Environmental factors that correlated with personality traits changed over developmental stages, suggesting that the environment influences age-specific personality (Stamps & Groothuis 2010). We saw boldness increase (in juveniles) and docility decrease (in yearlings and adults) as the day progressed. Although the exact reason for the within-day change is unknown, similar within-day changes in personality have also been seen in two species of coral reef fish due to temperature changes (Biro et al. 2010). Daily temperature changes may influence metabolic rate (Armitage 1991), which could influence personality traits (Biro et al. 2010). Body condition decreased boldness and docility in yearlings. These results are inconsistent with theoretical work suggesting that individuals in better body condition take more risks but incur fewer costs because they are better able to hold resources and escape predators (Luttbeg & Sih 2010).

In the three age classes, we found that predator presence increased boldness, as previously reported in fish (Brown et al. 2005). This intuitively makes sense, since individuals must strike a balance between being cautious in the presence of predators and maintaining potential foraging and reproductive opportunities (Cooper & Peréz-Mellado 2004). Individuals that are more cautious may therefore lose these opportunities in comparison to bolder individuals. Thus, we might expect bolder individuals in predator-rich areas despite the high risk.

Marmot interaction with humans seemed to have variable effects. Boldness increased with trial number and pedestrian traffic in yearlings and adults. Together, these factors suggest that marmots habituate to human disturbance over the summer (Li et al. 2011). Additionally, past research on eastern chipmunks, *Tamias striatus* (Martin & Réale 2008) and burrowing owls, *Athene cunicularia* (Carrete & Tella 2010) found a correlation between human disturbance and personality, suggesting habituation or habitat selection based on personality. In contrast, docility decreased in adults as pedestrian traffic increased. This coupled with an increase in docility as the days between capture increased suggest that marmots may become more sensitive to human interaction when trapped.

Yearlings that grew faster were more docile but also bolder. These are not consistent with the predictions made by Biro & Stamps (2008), who predicted that individuals that grow faster should engage in more risky behaviour to maintain that growth rate. Individual marmots do not have to protect or actively look for food patches (i.e. vegetation is abundant at our study site), thus decoupling the predicted link between risky behaviour and growth

Table 4
Repeatability (on diagonal), correlations (below diagonal) and pairwise comparison of repeatability (above diagonal) for docility and boldness for the different age classes (juvenile, yearling, adult) of yellow-bellied marmots

	Docility			Boldness		
	Juveniles	Yearlings	Adults	Juveniles	Yearlings	Adults
Juveniles	0.190 (0.021)***	NS	**	0.019 (0.027)	NS	NS
Yearlings	0.857 (0.159)***	0.268 (0.033)***	NS	0.379 (0.843)	0.373 (0.146)***	*
Adults	0.624 (0.161)***	0.782 (0.169)***	0.293 (0.039)***	−0.167 (−8.861)	0.446 (1.432)	0.053 (0.057)
Phenotypic	0.921 (0.029)	0.819 (0.038)	0.969 (0.041)	6.988 (4.665)	1.451 (0.235)	1.746 (0.228)

Estimates were obtained from trivariate models for each behaviour, considering each age class as a different trait. Estimates of repeatability are slightly different from univariate models (Table 1) because only significant fixed effects were used in these models to avoid overparametrization.
*$P<0.05$; **$P<0.01$; ***$P<0.001$; see text for exact *P* values (parameter bounded to zero).

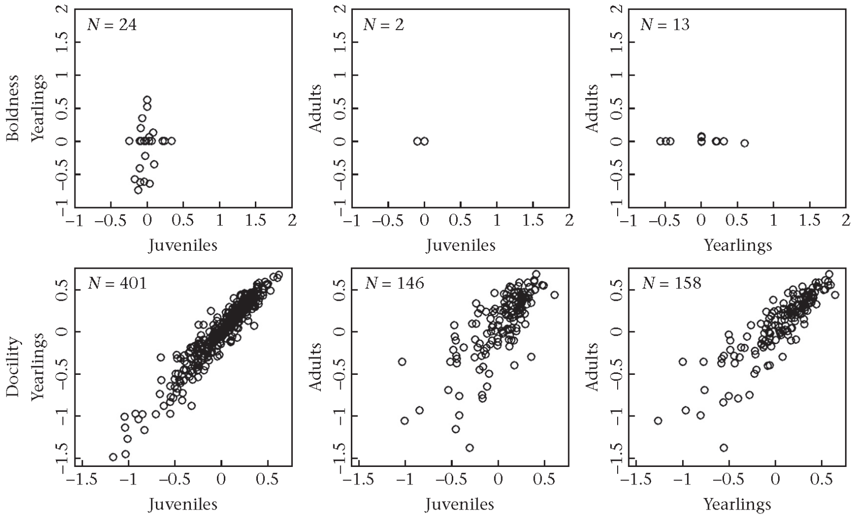


Figure 1. Relationships between juvenile–yearling, juvenile–adult and yearling–adult personality in yellow-bellied marmots. The top panel represents boldness (negative FID). The bottom panel represents docility (trapping behaviours). Best linear unbiased predictors (BLUPs) from trivariate models of boldness and docility are used for illustration purposes only. *N* represents the number of individuals for which data were available for both traits.

rate. In addition, faster-growing individuals might benefit from docility simply by not investing in more active and energetically costly nondocile behaviours.

Sex differences were found in yearling and adult boldness. Yearling males were bolder than females. All yearling males disperse, and this difference in boldness may be a method for individual males to prepare for dispersal. Alternatively, adult males were less bold than adult females. There is no apparent reason for this, but females have the added cost of gestation and lactation, and therefore need to be bolder to forage and survive hibernation (Andersen et al. 1976). Sex differences were also found in adult docility, with males being less docile than females. This result is inconsistent with the findings of a study of alpine marmots, *Marmota marmota*, where sexes did not differ in docility (Ferrari et al. 2013). However, Ferrari et al.'s (2013) study did not account for potential differences within each age group as our study did.

Repeatability within Life Stages

We found that boldness and docility developed differently. Boldness in juvenile and adult marmots was not repeatable; thus, it was not considered a personality trait in juveniles or adults. However, yearling marmots had significantly repeatable boldness levels. Interestingly, yearlings had higher repeatability than adults, suggesting that behaviour is not developmentally constrained. Docility, however, was repeatable in all age classes. These results suggest that these differences in personality development may allow for individuals to act adaptively at age-appropriate times.

Future research should focus on the potential reasons for a lack of boldness in juveniles and adults. Juveniles may lack individual differences because they have not yet undergone experiences that lead to differentiation (Freund et al. 2013). Alternatively, lack of a boldness personality trait might be due to differences in life-history strategy between each life stage. Body mass strongly affects overwinter survival in juveniles but does so to a lesser degree in adults (Lenihan & Vuren 1996). Juveniles triple their body mass in the 3 months from first emergence to hibernation. Thus, juveniles should prioritize foraging by adopting a single consistent strategy, while yearlings may adopt multiple strategies to cope with certain life stage events such as dispersal. Juveniles have also been found to be less vigilant than yearlings and adults in general, supporting the hypothesis that they prioritize foraging (Li et al. 2011). Juvenile, or smaller, poeciliid fish (*Brachyraphis episcopi*) were also found to be bolder than larger individuals, suggesting there is a trade-off between growth and potential mortality (Brown & Braithwaite 2004). Adults, however, must cope with a highly variable, harsh environment and must therefore be plastic (Armitage 1991).

Alternatively, docility was repeatable in all age classes, indicating that this personality trait develops early in life and canalizes with age. Individual differences in docility may therefore be linked to positive feedback loops with the environment. This result is supported by other studies showing that adults are less plastic than juveniles (Sinn et al. 2008; Gyuris et al. 2012). These differences in the development of personality traits are consistent with the hypothesis that consistent individual differences in behaviour can be adaptive and linked to life-history strategies (Wolf et al. 2007).

Stability within and between Traits across Time

We found differences between personality traits in stability across development. Boldness was not stable across development but docility was. Individuals' levels of boldness varied across all three life stages. However, the small sample of individuals with boldness data in multiple age classes limits conclusions from this data. We found no correlation between any age group. This result is consistent with the idea that behaviour is plastic and should change depending on the environment (Fox & Westneat 2010). Although the exact cause of the plasticity in personality is unknown, repeatability in boldness may be due to life stage events. For example, almost all yearling males and about half the yearling females disperse, and therefore, individuals exhibit different behavioural strategies to cope with dispersal or staying at the natal

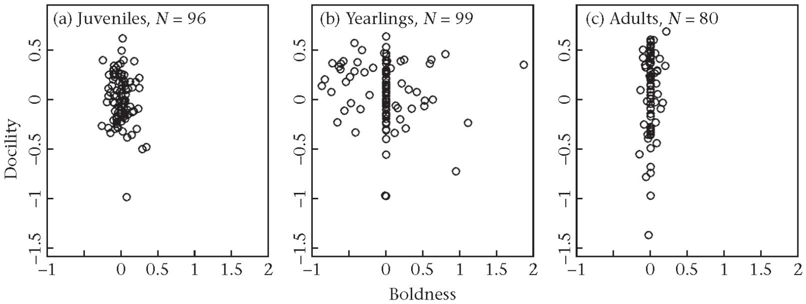


Figure 2. Relationship between docility and boldness in yellow-bellied marmot (a) juveniles, (b) yearlings and (c) adults. Best linear unbiased predictors (BLUPs) from bivariate models of docility and boldness are used for illustration purposes only. *N* represents the number of individual for which data were available for both traits.

colony. Again, adults have settled at a location and must cope with the changing yearly environment (Armitage 1991). Individuals are therefore changing their behavioural plasticity depending upon age-specific life-history events.

In contrast to boldness, docility appears to be stable throughout an individual's lifetime, given the strong positive correlations between the three age classes in the present study. This suggests that docility is established at birth or early in life and is stable throughout an individual's lifetime. Adult docility may result from previous environmental constraints and selection during development (Sinn et al. 2008).

Behavioural Syndrome

Boldness and docility have previously been shown to form a behavioural syndrome (Réale et al. 2009), but we found no correlation between boldness and docility in any life stage (Fig. 2). This is intuitive for juveniles and adults that do not show consistent individual differences in boldness, but do show consistency in docility. Although a syndrome could form later in life, the manner in which each personality trait develops might preclude such a formation.

Conclusion

In conclusion, we found that boldness and docility developed differently across life stages. This has major implications for our understanding of personality and raises a large number of questions about what might explain that difference. Juveniles grow rapidly and do not mate, yearlings disperse, and adults reproduce. Specific constraints and life histories of each age class are likely the basis for the observed differences in the ontogeny of personality. Selection can vary across life stages (Schluter et al. 1991; McNamara et al. 2009) and, therefore, different behavioural traits may be differentially important in those stages. Although we did not test for selection in this study, future research should focus on quantifying these different selection pressures throughout development and the potential adaptive reasons for the differences in the development of personality traits.

Acknowledgments

M.B.P. was supported by a GAANN fellowship, a GK-12 fellowship, and the UCLA Department of Ecology and Evolutionary Biology. D.M. was a National Science Foundation (NSF) REU fellow supported by DBI 0731346. V.A. was partially supported by a RMBL scholarship. J.G.A.M. was supported by a FQRNT fellowship and by the NSF. D.T.B was supported by the National Geographic Society, UCLA (Faculty Senate and the Division of Life Sciences), a Rocky Mountain Biological Laboratory research fellowship, and by the NSF (IDBR-0754247 and DEB-1119660 to D.T.B., as well as DBI 0242960 and 0731346 to the Rocky Mountain Biological Laboratory). We thank all the marmoteers over the past decade for help collecting the data, and the Blumstein lab for comments on previous versions of the manuscript. We also thank Lynn Fairbanks, Peter Nonacs, and three anonymous referees for their constructive feedback on previous versions of the manuscript.

References

Andersen, D. C., Armitage, K. B. & Hoffmann, R. S. 1976. Socioecology of marmots: female reproductive strategies. *Ecology*, **57**, 552–560.

Armitage, K. B. 1982. Yellow-bellied marmot. In: *CRC Handbook of Census Methods for Terrestrial Vertebrates* (Ed. by D. E. Davis), pp. 148–149. Boca Raton, Florida: CRC Press.

Armitage, K. B. 1986. Individual differences in the behavior of juvenile yellow-bellied marmots. *Behavioral Ecology and Sociobiology*, **18**, 419–424.

Armitage, K. B. 1991. Social and population dynamics of yellow-bellied marmots: results from long-term research. *Annual Review of Ecology and Systematics*, **22**, 379–407.

Armitage, K. B. & Downhower, J. F. 1974. Demography of yellow-bellied marmont populations. *Ecology*, **55**, 1233–1245.

Armitage, K. B. & Van Vuren, D. H. 2003. Individual differences and reproductive success in yellow-bellied marmots. *Ethology, Ecology & Evolution*, **15**, 207–233.

Baayen, R. H., Davidson, D. J. & Bates, D. M. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, **59**, 390–412.

Bates, D., Maechler, M. & Bolker, B. 2012. *lme4: Linear mixed-effects models Using S4 Classes, 2011.* http://CRAN.R-project.org/package=lme4.Rpackage version 0.999375-42.

Bell, A. M. & Stamps, J. A. 2004. Development of behavioural differences between individuals and populations of sticklebacks, *Gasterosteus aculeatus*. *Animal Behaviour*, **68**, 1339–1348.

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

Biro, P. A. & Stamps, J. A. 2008. Are animal personality traits linked to life-history productivity? *Trends in Ecology & Evolution*, **23**, 361–368.

Biro, P. A., Beckmann, C. & Stamps, J. A. 2010. Small within-day increases in temperature affects boldness and alters personality in coral reef fish. *Proceedings of the Royal Society B*, **277**, 71–77.

Blumstein, D. T. 2003. Flight-initiation distance in birds is dependent on intruder starting distance. *Journal of Wildlife Management*, **67**, 852–857.

Blumstein, D. T. 2010. Flush early and avoid the rush: a general rule of antipredator behavior? *Behavioral Ecology*, **21**, 440–442.

Blumstein, D. T., Wey, T. W. & Tang, K. 2009. A test of the social cohesion hypothesis: interactive female marmots remain at home. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 3007–3012.

Brown, C. & Braithwaite, V. A. 2004. Size matters: a test of boldness in eight populations of the poeciliid *Brachyraphis episcopi*. *Animal Behaviour*, **68**, 1325–1329.

Brown, C., Jones, F. & Braithwaite, V. 2005. In situ examination of boldness–shyness traits in the tropical poeciliid, *Brachyraphis episcopi*. *Animal Behaviour*, **70**, 1003–1009.

Brydges, N. M., Colegrave, N., Heathcote, R. J. & Braithwaite, V. A. 2008. Habitat stability and predation pressure affect temperament behaviours in populations of three-spined sticklebacks. *Journal of Animal Ecology*, **77**, 229–235.

Careau, V., Rêale, D., Humphries, M. M. & Thomas, D. W. 2010. The pace of life under artificial selection: personality, energy expenditure, and longevity are correlated in domestic dogs. *American Naturalist*, **175**, 753–758.

Carrete, M. & Tella, J. L. 2010. Individual consistency in flight initiation distances in burrowing owls: a new hypothesis on disturbance-induced habitat selection. *Biology Letters*, **6**, 167–170.

Cooper, W. E., Jr. 2009. Variation in escape behavior among individuals of the striped plateau lizard *Sceloporus virgatus* may reflect differences in boldness. *Journal of Herpetology*, **43**, 495–502.

Cooper, W. E., Jr. & Perêz-Mellado, V. 2004. Tradeoffs between escape behavior and foraging opportunity by the Balearic lizard (*Podarcis lilfordi*). *Herpetologica*, **60**, 321–324.

Dill, L. M. & Houtman, R. 1989. The influence of distance to refuge on flight initiation distance in the gray squirrel (*Sciurus carolinensis*). *Canadian Journal of Zoology*, **67**, 233–235.

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

Dingemanse, N. J., Both, C., Drent, P. J. & Tinbergen, J. M. 2004. Fitness consequences of avian personalities in a fluctuating environment. *Proceedings of the Royal Society B*, **271**, 847–852.

DiRienzo, N., Pruitt, J. N. & Hedrick, A. V. 2012. Juvenile exposure to acoustic sexual signals from conspecifics alters growth trajectory and an adult personality trait. *Animal Behaviour*, **84**, 861–868.

Ferrari, C., Pasquaretta, C., Carere, C., Cavallone, E., von Hardenberg, A. & Rêale, D. 2013. Testing for the presence of coping styles in a wild mammal. *Animal Behaviour*, **85**, 1385–1396, http://dx.doi.org/10.1016/j.anbehav.2013.03.030.

Fox, C. W. & Westneat, D. F. 2010. Adaptation. In: *Evolutionary Behavioral Ecology* (Ed. by D. F. Westneat & C. W. Fox), pp. 16–31. New York: Oxford University Press.

Frase, B. A. & Hoffmann, R. S. 1980. Marmota flaviventris. *Mammalian Species*, **135**, 1–8.

Freund, J., Brandmaier, A. M., Lewejohann, L., Kirste, I., Kritzler, M., Küger, A., Sachser, N., Lindenberger, U. & Kempermann, G. 2013. Emergence of individuality in genetically identical mice. *Science*, **340**, 756–759.

Gilmour, A. R., Gogel, B., Cullis, B., Thompson, R., Butler, D., Cherry, M., Collins, D., Dutkowski, G., Harding, S. & Haskard, K. 2009. *ASReml User Guide.* Release 3.0. Hemel Hempstead: VSN International.

Groothuis, T. G. G., Carere, C., Lipar, J., Drent, P. J. & Schwabl, H. 2008. Selection on personality in a songbird affects maternal hormone levels tuned to its effect on timing of reproduction. *Biology Letters*, **4**, 465–467.

Gyuris, E., Feró, O. & Barta, Z. 2012. Personality traits across ontogeny in firebugs, *Pyrrhocoris apterus*. *Animal Behaviour*, **84**, 103–109.

Lenihan, C. & Vuren, D. V. 1996. Growth and survival of juvenile yellow-bellied marmots (*Marmota flaviventris*). *Canadian Journal of Zoology*, **74**, 297–302.

Li, C., Monclûs, R., Maul, T. L., Jiang, Z. & Blumstein, D. T. 2011. Quantifying human disturbance on antipredator behavior and flush initiation distance in yellow-bellied marmots. *Applied Animal Behaviour Science*, **129**, 146–152.

Luttbeg, B. & Sih, A. 2010. Risk, resources and state-dependent adaptive behavioural syndromes. *Philosophical Transactions of the Royal Society B*, **365**, 3977–3990.

McNamara, J. M., Houston, A. I., Barta, Z., Scheuerlein, A. & Fromhage, L. 2009. Deterioration, death and the evolution of reproductive restraint in late life. *Proceedings of the Royal Society B*, **276**, 4061–4066.

Martin, J. G. & Rêale, D. 2008. Animal temperament and human disturbance: implications for the response of wildlife to tourism. *Behavioural Processes*, **77**, 66–72.

Pinheiro, J. C. & Bates, D. M. 2000. *Mixed-effects Models in S and S-Plus*. New York: Springer Verlag.

Pruitt, J. N., Cote, J. & Ferrari, M. C. O. 2012. Behavioural trait variants in a habitat-forming species dictate the nature of its interactions with and among heterospecifics. *Functional Ecology*, **26**, 29–36.

R Development Core Team 2011. *R: a Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

Rêale, D., Gallant, B. Y., Leblanc, M. & Festa-Bianchet, M. 2000. Consistency of temperament in bighorn ewes and correlates with behaviour and life history. *Animal Behaviour*, **60**, 589–597.

Rêale, D., Reader, S. M., Sol, D., McDougall, P. T. & Dingemanse, N. J. 2007. Integrating animal temperament within ecology and evolution. *Biological Reviews*, **82**, 291–318.

Rêale, D., Martin, J., Coltman, D. W., Poissant, J. & Festa-Bianchet, M. 2009. Male personality, life-history strategies and reproductive success in a promiscuous mammal. *Journal of Evolutionary Biology*, **22**, 1599–1607.

Rödel, H. G. & Meyer, S. 2011. Early development influences ontogeny of personality types in young laboratory rats. *Developmental Psychobiology*, **53**, 601–613.

Runyan, A. M. & Blumstein, D. T. 2004. Do individual differences influence flight initiation distance? *Journal of Wildlife Management*, **68**, 1124–1129.

Schluter, D., Price, T. D. & Rowe, L. 1991. Conflicting selection pressures and life history trade-offs. *Proceedings of the Royal Society B*, **246**, 11–17.

Sinn, D. L., Gosling, S. D. & Moltschaniwskyj, N. A. 2008. Development of shy/bold behaviour in squid: context-specific phenotypes associated with developmental plasticity. *Animal Behaviour*, **75**, 433–442.

Stamps, J. 2007. Growth-mortality tradeoffs and ‘personality traits’ in animals. *Ecology Letters*, **10**, 355–363.

Stamps, J. & Groothuis, T. G. G. 2010. The development of animal personality: relevance, concepts and perspectives. *Biological Reviews*, **85**, 301–325.

Svendson, G. E. 1974. Behavioral and environmental factors in the spatial distribution and population dynamics of a yellow-bellied marmot population. *Ecology*, **55**, 760–771.

Svendson, G. E. & Armitage, K. B. 1973. Mirror-image stimulation applied to field behavioral studies. *Ecology*, **54**, 623–627.

Wilson, A. D. M. & Krause, J. 2012. Personality and metamorphosis: is behavioral variation consistent across ontogenetic niche shifts? *Behavioral Ecology*, **23**, 1316–1323.

Wolf, M., van Doorn, G. S., Leimar, O. & Weissing, F. J. 2007. Life-history trade-offs favour the evolution of animal personalities. *Nature*, **447**, 581–584.

Ydenberg, R. & Dill, L. 1986. The economics of fleeing from predators. *Advances in the Study of Behavior*, **16**, 229–249.

CHAPTER 4

MAINTENANCE OF RISKY PERSONALITY TRAITS IN YELLOW-BELLIED MARMOTS

Summary

1. Individuals vary consistently in their risk-taking behaviour and this can be considered a personality trait. Several hypotheses for the causes and maintenance of differences in risky personality traits have been suggested to explain why individuals are stable in their behaviour. First, personality is associated with growth or productivity. Individuals that grow more quickly will take more risks to maintain that growth. Second, personality is linked to residual reproductive value. Individuals with higher reproductive residual value, or young individuals, should be more risk-averse than older, lower valued, individuals. Finally, personality is maintained by body condition. Individuals in better body condition are able to defend against predators and are thus more risk-prone.
2. We simultaneously evaluated these three hypotheses to investigate what maintains variation in three risk related traits, docility, activity and exploration, in yellow-bellied marmots (*Marmota flaviventris*).
3. We found no evidence of growth, residual reproductive value, or body condition effects explaining between individual variation in docility, activity nor exploration. However, we did find that other fixed effects differentially influenced personality traits between contexts.
4. The maintenance of risky personality traits in yellow-bellied marmots thus remains enigmatic. As more investigators evaluate multiple hypotheses, we will learn more about what maintains variation in personalities.

1. Introduction

Our understanding of animal personalities, or consistent individual differences in behaviour across time and context, has advanced greatly in recent years (Sih *et al.*, 2004; Réale *et al.*, 2007; Pruitt & Ferrari, 2011). Despite recent advances in the ecological and evolutionary consequences of personality (Sih *et al.*, 2004; Réale *et al.*, 2007; Smith & Blumstein, 2008), little empirical work has been done on the causes and maintenance of intra-individual differences. A number of theoretical explanations have been presented that link state and state-dependent behaviour to explain the maintenance of personality (Stamps, 2007; Wolf *et al.*, 2007; Biro & Stamps, 2008; Dingemanse & Wolf, 2010; Luttbeg & Sih, 2010; Wolf & Weissing, 2010). The three most common hypotheses are that differences in growth rate, residual reproductive value, or body condition can explain the differences and maintenance of personality.

First, individuals may vary in life history strategies and this variation offers a potential explanation for the maintenance of personality. For example, differences in productivity, specifically in growth rate, metabolism, or fecundity, could drive personality trait differences (Stamps, 2007; Biro & Stamps, 2008) whereby individuals with higher productivity take more risks (e.g. they are bolder, more aggressive, more exploratory) to maintain initial levels of productivity. Individuals maintain this productivity because physiological and life-history trajectories are set early in life and it is costly to deviate from a given trajectory (Biro & Stamps, 2008). Empirical evidence for the productivity hypothesis is unclear with mixed support in different taxa (Sundström et al. 2004; Ward et al. 2004; Biro & Post 2008; Careau et al. 2009; Edenbrow & Croft 2011).

Second, differences in an individual’s residual reproductive value, coupled with state-dependent behaviour, have the potential to maintain personality differences (Wolf *et al.*, 2007).

Because individuals may incur a trade-off between reproduction early or late in life, individuals with higher future reproductive value should take fewer risks so as to protect future assets (Clark, 1994). Empirical evidence for residual reproductive value correlating with personality has been shown in killifish (*Kryptolebias marmoratus*) (Edenbrow & Croft, 2011), grey mouse lemurs (*Microcebus murinus*) (Dammhahn, 2012), and great tits (*Parus major*) (Nicolaus *et al.*, 2012).

Third, differences in body condition, size, or energy reserves may maintain variation in personality (Luttbeg & Sih, 2010). Individuals in better condition are better able to avoid predators or fight conspecifics. Thus, body condition is maintained because individuals with higher body condition are better able to access or defend resources, and escape predators, creating a positive feedback (i.e. state-dependent safety). Individuals with lower body condition are unable to take risks because they do not react accordingly (escape predators or fight effectively) and are incapable of increasing their condition (e.g. making the best of a bad job). However, empirical evidence for this hypothesis is equivocal (Martin & Réale, 2008).

Hypotheses to maintain personality are not mutually exclusive, thus, it is important to test them simultaneously to understand the relative importance of each. Previous studies, however, typically test only one of the hypotheses suggested to maintain personality (Sundström *et al.*, 2004; Ward *et al.*, 2004; Biro & Post, 2008; Martin & Réale, 2008; Careau *et al.*, 2009; Edenbrow & Croft, 2011; Vainikka *et al.*, 2011; Dammhahn, 2012; Nicolaus *et al.*, 2012), and none, to our knowledge, have explicitly used a multiple hypothesis approach in a wild population. Our goal was to simultaneously evaluate the three most common hypotheses (productivity, residual reproduction and condition) in a wild population. We capitalised on a long-term study of yellow-bellied marmots (*Marmota flaviventris*) to study factors that explain the maintenance of docility, activity, and exploration—all of which may reflect risky personality traits. This system offers a great opportunity to use a multiple hypothesis approach; individuals differ in both the rate of mass gain, body condition, and they live up to 16 years (Armitage & Downhower, 1974) which allows the study of the correlates of variation in potential future reproduction.

Each of the three hypotheses makes specific predictions that can be used to disentangle whether one or all mechanisms are associated with maintenance of personality. If risk-prone behavioural traits are maintained by differences in productivity (Biro & Stamps, 2008), we expect that individuals that grow faster will be bolder to maintain their growth trajectories. If these traits are maintained by residual reproductive value, we expect there to be personality differences by age (Stamps, 2007). Newly reproductive mature individuals have high residual reproductive value and should therefore be risk averse. Very young and very old individuals have lower residual reproductive value and are thus predicted to be more risky (Pianka & Parker, 1975). Finally, if body condition maintains risk prone personality traits, we expect individuals with greater mass to take more risks (Luttbeg & Sih, 2010). It should be noted that maintenance of personality will be reflected between individuals but that those mechanism could also explain within individual variation in behaviours. Consequently, we evaluated the three hypotheses at both between (i.e. personality maintenance) and within individual level using an individual centering approach (Dingemanse & Dochtermann, 2013).

2. Material and methods

(a) Subjects

We studied yellow-bellied marmots in and around the Rocky Mountain Biological Station, Gothic, Co, USA from 2002-2012. Marmots are large, facultatively social, semi-fossorial sciurid rodents that live in colonies. Colonies consist of matrilineal groups made of a mother and her daughters or sisters and their female offspring while adult males guard and patrol territories (Frase & Hoffmann, 1980). Marmots are of known age because they are first trapped as juveniles. Age classes in marmots are defined as: juveniles, or young of the year; yearlings, or individuals that survived the first winter; and adults, or individuals two years or older (Armitage & Downhower, 1974). We trapped individuals regularly between mid-April to mid-September using Tomahawk-live traps placed at the entrances of burrows. Individuals were transferred to a cloth-handling bag, ear-tagged, weighed, sexed, reproductive status checked, and given a unique fur mark (Blumstein *et al.*, 2009).

(b) Quantifying personality

We quantified docility from 8990 trapping events on 1201 individuals. Docility is defined here as an individual's reaction to being trapped and handled (Réale *et al.*, 2000) and may describe risk-taking behaviour (Careau *et al.*, 2009). At each trapping event, we noted whether individuals struggled in the trap, tooth-chattered, bit the cage, alarm called, and hesitated to walk into the handling bag. We dichotomously (0/1) scored these behaviours, summed them, and subtracted this from the total potential score to attain a docility index for that trapping event. Individuals that scored 0 are non-docile and can also be considered defiant while individuals that scored 5 are docile and easily handled (Réale *et al.*, 2000).

We quantified exploration/activity using Open field (OF) and Mirror Image Stimulation (MIS) tests. From 2010-2012, we conducted 614 Open field and Mirror Image Stimulation trials on 226 individuals. Individuals were transferred from the trap to a cloth-handling bag, their heart rate was measured for 15 s, and individuals were then transported to an arena measuring 91.4

cm^3 made of 0.47 cm opaque PVC sheeting. A mirror (30.5 cm x 61.0 cm) was placed at the base of one side of the arena and covered with an opaque sliding door. A door (61.0 cm^2) was cut out of the side opposite the mirror. A 16-square grid (~22.85 cm^2) was drawn on the floor of the arena to determine activity and exploration. All trials were video-recorded from above for later behavioural scoring. Individuals were released at the trap site location post-trial. Each trial consisted of a three minute OF test where individuals could freely move and explore the arena. Immediately following the OF test, the sliding door was removed to expose the mirror for an additional three minutes. The MIS test was similar to the OF test with the addition of the mirror. Individuals were tested a maximum of once a day. Technical issues (e.g. camera failure) along with inability to measure certain aspects of the individual including predator pressure or mass at capture restricted our OF analysis to 435 trials on 178 individuals and restricted our MIS analysis to 428 trials on 177 individuals.

All trials were scored with JWatcher 1.0 (Blumstein & Daniel, 2007). We calculated number of events and proportion of time alarm calling, jumping, looking, walking, and sniffing the arena. In addition, we also calculated grid lines crossed and proportion of squares visited for both OF and MIS, and for MIS, the proportion of time scratching or pawing the mirror, and the total proportion of time at the mirror. All scorers were trained to have >95% inter- and intra-observer agreement. Raw scores for both tests were subjected to a Principle Components Analysis (PCA) and the resulting factor scores for each trial were used for further analysis. We used SPSS v. 18.0 (Chicago, IL) with varimax rotation for both PCAs.

(c) Quantifying key environmental and biological variation

To understand how the environment might influence personality, we quantified pedestrian traffic for colonies and predator presence. Pedestrian traffic was calculated during the peak tourist season in the summer of 2010 by counting the number of pedestrians travelling within 300 m of six colonies (Li *et al.*, 2011). Predator presence was calculated using the number of predators seen at each colony prior to June each year, and we divided this number by the number of observations at that colony. Full methods can be found in (Petelle *et al.*, 2013). Mass gained between 15 June and 15 August (hereafter growth rate) was estimated by taking the difference of estimated body mass on 15 June and 15 August. Body mass on these dates was estimated by extracting best linear unbiased predictors using a linear mixed effects model approach that had a quadratic function of day of the year (see supplementary material) (Martin & Pelletier, 2011) Petelle MB & Blumstein DT, unpublished data).

(d) Statistical analysis

First, to estimate the repeatability of the different behaviours, we fitted a linear mixed effects model (LMM) for docility, activity, and exploration without the three variables related to the personality maintenance hypotheses we were testing (Table 1a). We then fit the same model with the addition of growth rate (index of productivity), a quadratic function of age (index of residual reproductive value), and mass at capture (index of body condition) as fixed effects (Table 1b). We used an individual centering approach (individual mean and deviation from the individual mean) for both growth and mass (Dingemanse & Dochtermann, 2013). This approach also allows us to estimate the between (individual mean) and within (deviation from individual mean) individual effects of growth and mass on behaviours and assist with model convergence

(Dingemanse & Dochtermann, 2013). A significant effect of the individual mean is related to a decrease in between individual variance (compared to a model without the effect) and is interpreted as explaining, at least partly, personality. An effect of the deviation from the individual mean would be related to a decrease in the residual variance and would be interpreted as explaining within individual variation, but not personality. Since individuals were observed at different ages and for a different number of years, it was not possible to use an individual centering approach with age. To correct for differences between masses among different age classes (juvenile, yearling, and adult) we calculated individual means within each age class.

Other independent variables included time (AM/PM), date, pedestrian traffic, predator presence (for activity and exploration only), sex, trial number, and days between trials. The interaction was removed from the model if it was non-significant ($p > 0.05$). We included these factors because they have been shown to influence personality in a previous study on docility (Petelle *et al.*, 2013). Individual identity and year were included as random effects. We used a log-likelihood ratio test to determine significance of random effects (Pinheiro & Bates, 2000). A significant effect of identity indicates a personality trait. Repeatability was estimated by taking the individual variance divided by the total phenotypic variance after accounting for fixed effects. All models were fitted in R 2.14 (RDevelopment, 2012) with the package lme4 (Bates *et al.*, 2012). We calculated MCMC *p*-values using the pvals.fnc function in the language R package (Baayen *et al.*, 2008). Metadata are archived at www.eeb.ucla.edu/Faculty/Blumstein/MarmotsOfRMBL/data.html

3. Results

(a) PCA and repeatability

Principle component analysis of Open Field and Mirror Image Stimulation data resulted in the extraction of four and six factors, respectively. The first OF and MIS components were characterized by activity, while the second OF and third MIS component were labelled as exploratory (see Supplementary Table 1 for component loadings). We thus limited our analyses to the first and second OF components and the first and third MIS factors.

Docility was significantly repeatable. Both activity and exploration were repeatable in both OF and MIS tests (Table 1a and Table 1b). We can therefore consider docility, activity, and exploration to be personality traits within this population.

(b) Maintenance of personality

Productivity: We found no significant effects of individual mean (between individual) growth rate on either docility (Table 2; Figure 1) or activity or exploration in OF and MIS (Table 2, Figure 2). Additionally, deviation from the individual mean (within individual) in growth rate did not significantly influence docility, activity, or exploration (Table 2). **Residual Reproductive Value:** We found no significant effect of age on any personality trait, nor was there a within-individual effect of age on personality (Table 2; Figure 1-3). **Body Condition:** Between individual differences (mean individual effects) in mass were not associated with any personality trait (Table 2: Figures 1-3). Interestingly, within individual effects (deviation from the individual mean in mass) was significant for docility and OF activity (Table 2). As expected given their non-significance, inclusion of the three hypotheses in models of personality traits did not change variance components (i.e. repeatability, Table 1b).

(c) Environmental factors

In addition to the three hypotheses we were testing, we found that individuals were less docile and more exploratory in the OF test as the active season progressed (docility: $\beta \pm SE = -1.21e-03 \pm 6.99e-04$, $p < 0.038$; OF: $8.25e-03 \pm 3.56e-03$, $p = 0.012$). Individuals were less docile the later in the day (-0.144 ± 0.028, $p < 0.001$). Pedestrian traffic was associated with more exploration in the OF ($0.011 \pm 7.60e-03$, $p = 0.026$). With repeated tests, individuals were more docile ($0.030 \pm 4.95e-03$, $p < 0.001$) and less active in the OF test (-0.156 ± 0.048, $p < 0.001$). The longer the period between trials, individuals became more docile ($2.35e-03 \pm 1.14e-03$, $p < 0.038$) and less active in OF tests (-0.016 ± 0.005, $p < 0.001$). We also found that individuals in areas with higher predator presence were more active in both the OF (3.710 ± 0.761, $p < 0.001$) and in the MIS (2.130 ± 0.976, $p = 0.012$) tests. An increase in pedestrian traffic had a corresponding increase in exploration in the OF test ($0.001 \pm 7.60e-03$, $p = 0.026$) and decrease in activity in the MIS test ($-0.023 \pm 8.09e-03$, $p = 0.016$).

4. Discussion

Despite having large sample sizes, and sufficient power to detect other significant effects, our investigation into the causes and maintenance of risky personality revealed no support for any of the three hypotheses tested (productivity, residual reproduction and body condition) on any of the five behavioural measures representing three different personality traits (docility, activity and exploration). In addition, we found that different environmental variables affect the same personality trait in different contexts or tests (OF and MIS).

Environmental Variables

Although not the main focus of this study, other fixed effects included in these models to correct for potential bias in the data (Supplementary Table 2) offer insights into how personality traits are influenced by factors within and between contexts. These between context effects were most prominently shown in activity between the OF and MIS tests. We found that activity was influenced by days between trial and predator presence similarly in both the OF and MIS trials. Activity, however, significantly decreased with trial suggesting habituation (Rankin *et al.*, 2009) in the OF test only. The absence of habituation in MIS might be due to the nature of the test and the stimulation by the mirror or it might be an experimental bias. How activity is associated with sociability (a commonly measured trait in MIS tests (Svendsen & Armitage, 1973)) is unknown in this population. MIS tests were done by showing a mirror after 3 minutes of the OF test. If activity decreases with time spent in the arena, activity might have already reached a low point by the start of the MIS test, and thus showing no habituation. It should be noted that because activity was estimate through PCA components estimated separately for OF and MIS, we were not able to compare the mean activity level between tests. We also found that activity significantly decreased with higher pedestrian traffic in the MIS test but not in the OF test. This suggests that anthropogenic disturbance has a strong impact on behaviours in a social context (Duchesne *et al.*, 2000; Lacy & Martins, 2003). Exploration in the OF test was associated with date and pedestrian traffic. These effects were not seen in the MIS test. Because exploration may have occurred mostly during the OF test, these effects may have been diminished during the MIS test. Indeed, we saw none of the effects included in the MIS model significantly influencing exploration. This is not to say that there may be differences in the effects between contexts. Both days between trial and time of day change sign between contexts.

In addition to activity and exploration, we found that individuals were less docile later in the active season, with repeated trials, and as the day progressed (Biro *et al.*, 2010; Petelle *et al.*, 2013). Small, within day changes in personality has been reported previously in this system (Petelle *et al.*, 2013). These changes may be linked to temperature and within day metabolic changes (Biro *et al.*, 2010).

Interestingly, we find that personality traits are potentially influenced by different factors between contexts (OF and MIS tests). We suggest that studies investigate personality traits between contexts to better understand how their environment is influencing these traits. How the environment influences the magnitude and direction of selection on traits in different situation gives us a more comprehensive of personality in the wild, and may help us predict how selection may shift personality traits in a changing environment.

Maintenance of personality

We found no evidence that differences in growth are associated with differences in any personality trait. Our previous research shows that faster growing yearlings are more docile (Petelle *et al.*, 2013), but this effect disappeared when all ages were included. Growth rate and other metabolic traits or life history traits, considered part of the pace of life syndrome (Réale *et al.*, 2010), have been found to be associated with a number of personality traits. Indeed, a number of studies have found a positive link between growth and risk taking (Clobert *et al.*, 2000; Ward *et al.*, 2004; Pottinger, 2006; Careau *et al.*, 2008). However, in one of the largest studies of a free-living population, Bouwhuis et al. (2014), found a weak correlation between basal metabolic rate and personality in great tits (*Parus major*).

We also found no evidence that residual reproductive value is likely to cause and maintain personality variation. Our results suggest that neither very young nor very old individuals are more apt to take risks while newly reproductively mature adults are more cautious. However, variation in reproductive strategies both within and between sexes might affect our results. Long-winged firebugs (*Pyrrhocoris apterus*) are supposed to have lower residual reproductive value than the short winged ones are more risky (Gyuris *et al.*, 2011). This effect however was observed only in females and is explained by differences in reproductive investment both between sexes and morphs. In one of the largest studies of life history traits and personality, older bighorn rams (*Ovis canadensis*) that were bolder had higher reproductive success. However, the study also found that more docile, older individuals also had higher reproductive success (Réale *et al.*, 2009) indicating that within a sex, differences in reproductive strategies could lead to different relationship between personality, age and reproduction. We found no such effect of sex in our analysis, although this may be due to the underrepresentation of adult males in our study ($N_{docility} = 67$; $N_{OF/MIS} = 17$). In addition a better understanding of variation in reproductive strategies of both sexes is needed to clarify its impact of behavioural variation and personality maintenance.

We found no association between differences in mass and differences in risk-taking behaviour. Previous work on this species found that heavier yearlings were less docile (Petelle *et al.*, 2013), but this effect was not seen when we included all age groups in the analysis. No association between body mass and docility was observed in bighorn ewes (Réale *et al.*, 2000), however, heavier eastern chipmunks (*Tamius striatus*) were found to be less docile (Martin & Réale, 2008). Thus, our results are inconsistent with the safety-dependent hypothesis, which predicts that individuals in better body condition should take more risks. Differences in mass did not have an effect on activity or exploration in either the OF or MIS test. A previous review of muroid rodent personality found no association between OF exploration and mass (Careau *et al.*, 2009). Additionally, MIS tests are typically used to assess social attributes, and body condition appears not to influence activity within this social context. However, previous research has shown that body condition does influence both the probability of dispersal as well as an individual's social rank (Huang *et al.*, 2011). Whether there is a life-long link between activity, exploration and dispersal remains to be determined in this species. Risk-prone behaviour has been shown to be part of a dispersal syndrome, but age-dependent variation of boldness was found to be related to the age at which animal disperse (Petelle *et al.*, 2013).

Interestingly, individuals that deviated from their mean body mass varied their docility and activity level. Individuals that were lighter than their average had higher levels of docility. We are unsure why this might be, but when an individual is lighter they may be conserving energy by not struggling when trapped. However, when lighter than on average individuals compensated by being more active and inversely were less active when bigger. This makes intuitive sense as individuals are attempting to gain more mass they may need to be more active or explorative to find food. Likewise, mass compensation influenced activity and exploratory behaviour in zebra finches (*Taeniopygia guttata*) (Krause & Naguib, 2011). However, in that study, individuals with higher compensatory growth were less explorative and less active.

In this study, we assume that the OF and MIS components reflect the same personality traits because they have similar component loadings in a PCA. Thus, activity and exploration were measured and evaluated in two different contexts, OF and MIS. We found that different environmental variables were significant between analyses of activity and exploration in OF and MIS. This could suggest that selection may act upon a personality trait through different environmental variables between contexts (Sinervo & Calsbeek, 2010). A selection analysis between these contexts is beyond the scope of this paper. However, it could also suggest that despite their similarities in experimental set-up and component loadings of behaviour in a PCA, OF and MIS tests are capturing different personality traits that are not directly equivalent.

While significant covariates are able to explain variation in personality traits, none of the predictions from three proposed hypotheses (productivity, residual reproductive value, or body condition) explain personality variation in this marmot population. Although there are multiple studies that investigate one or more of these hypotheses, this is the first to explicitly test simultaneously the three most common on a large, multi-year dataset from a free-living population. There are other potential mechanisms that may explain variation in personality including other physiological traits or social niche specialisation (Bergmüller & Taborsky, 2010; Laskowski & Pruitt, 2014). Future studies should specifically evaluate other potential mechanisms.

Acknowledgements. We thank the many marmoteers who helped us collect trapping data since 2001. We would also like to thank Lynn Fairbanks and Peter Nonacs for comments and discussion on a previous draft. Research was conducted under UCLA OPRS permit 2001-191.

Data accessibility. Data are archived at www.eeb.ucla.edu/Faculty/Blumstein/MarmotsOfRMBL/data.html.

Funding statement. M.B.P. was supported by a U.S. Department of Education GAANN Fellowship, a NSF GK12 Fellowship, and the University of California Los Angeles. J.G.A.M. was supported by a Marie-Curie International Incoming Fellowship. D.T.B was supported by the

National Geographic Society, UCLA (Faculty Senate and the Division of Life Sciences), a Rocky Mountain Biological Laboratory research fellowship, and by the NSF (IDBR-0754247 and DEB-1119660 to D.T.B., as well as DBI 0242960 and 0731346 to the Rocky Mountain Biological Laboratory).

Table 4-1.

Without growth, age class, and mass (1a)			Variance			Ratio		LRT		P-value	
Trait	N obs-ind	Mean (SD)	Identity	Year	Pheno-typic	Identity	Year	Identity	Year	Identity	Year
Docility	6814-941	4.289 (0.932)	0.178	0.011	0.866	0.206	0.013	743.095	47.869	<0.001	<0.001
Activity (OF 1)	398-166	0 (1)	0.142	0.022	0.701	0.203	0.031	9.328	2.443	0.002	0.118
Exploration (OF 2)	398-166	0 (1)	0.350	0.157	1.189	0.295	0.132	15.755	16.562	<0.001	<0.001
Activity (MIS 1)	392-165	0 (1)	0.543	0.052	1.100	0.495	0.048	24.564	5.314	<0.001	0.020
Exploration (MIS 3)	392-165	0 (1)	0.016	0.001	0.060	0.265	0.021	8.844	1.158	0.003	0.282
With growth, age class, and mass (1b)			Variance			Ratio		LRT		P-value	
Trait	N obs-ind	Mean (SD)	Identity	Year	Pheno-typic	Identity	Year	Identity	Year	Identity	Year
Docility	6814-941	4.289 (0.932)	0.179	0.011	0.861	0.208	0.013	749.906	48.975	<0.001	<0.001
Activity (OF 1)	398-166	0 (1)	0.158	0.016	0.649	0.244	0.025	17.061	0.875	<0.001	0.350

Exploration (OF 2)	398-166	0 (1)	0.333	0.127	1.135	0.294	0.112	14.980	12.308	<0.001	<0.001
Activity (MIS 1)	392-165	0 (1)	0.573	0.050	1.110	0.517	0.045	30.858	3.458	<0.001	0.063
Exploration (MIS 3)	392-165	0 (1)	0.014	9.05e-04	0.059	0.244	0.015	7.413	0.601	0.006	0.438

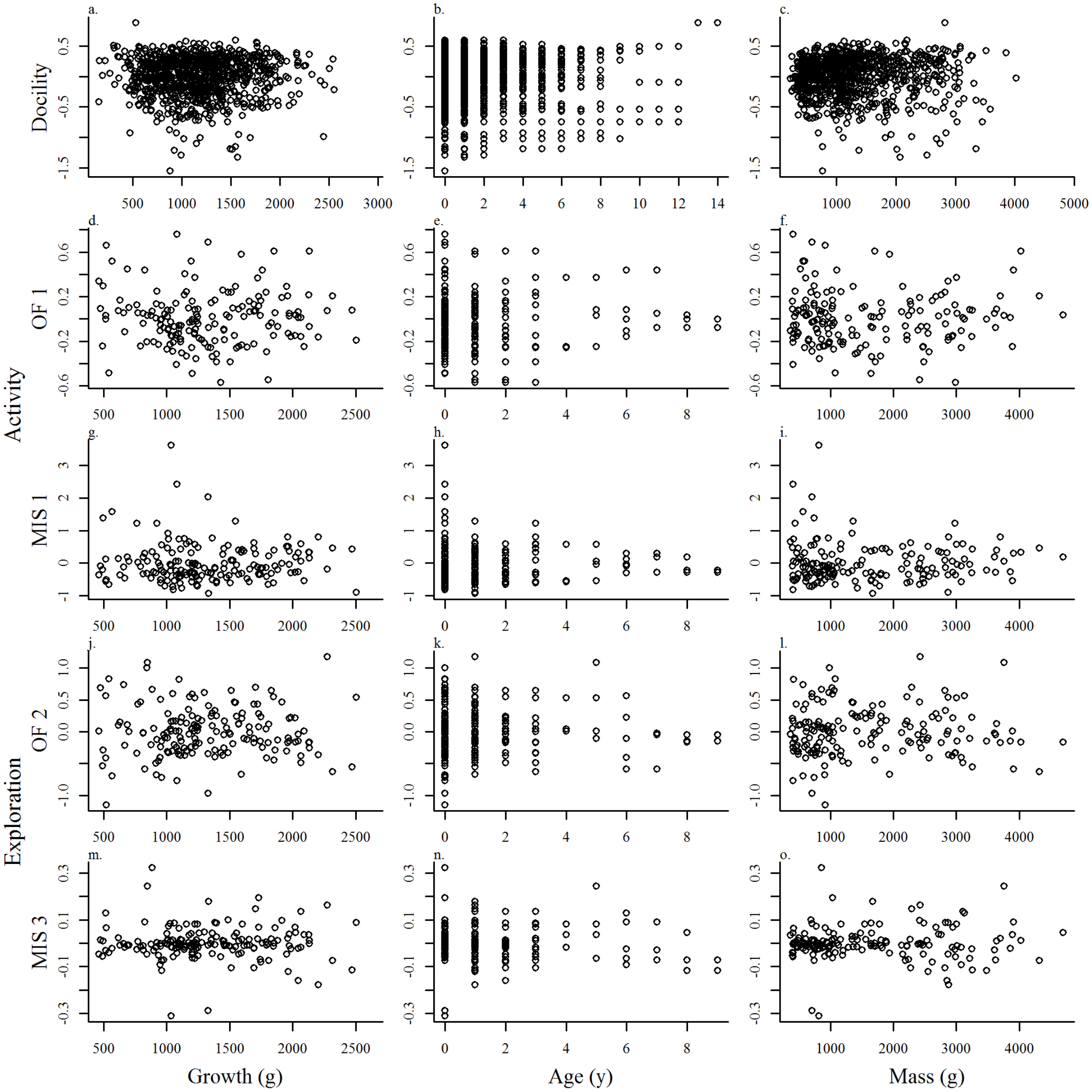
Variance components, variance ratio, log-likelihood ratio tests (LRT) and significance of random effects with and without growth, age, and mass for docility (in the trap) and activity and exploration during open field (OF) and mirror image stimulation (MIS) tests. Table 1a reports information for models without growth rate, age class, and mass as fixed effects, while Table 1b describes the models with them. Activity and exploration in OF and MIS were estimated from principal component analyses (factors 1 and 2 for OF, factor 1 and 3 for MIS respectively) and have a mean of 0 and a standard deviation of one. Total number of individuals, trials, mean of the traits, and standard deviation are also reported. We added four and square root transformed Exploration (MIS 3) to normalize residuals.

Table 4-2.

	Docility			Activity (OF)			Exploration (OF)			Activity (MIS)			Exploration (MIS)		
	Estimate	S.E.	P-value	Estimate	S.E.	P-value	Estimate	S.E.	P-value	Estimate	S.E.	P-value	Estimate	S.E.	P-value
Growth (between)	4.00 e-05	4.85 e-05	0.324	1.55 e-04	1.67 e-04	0.183	2.97 e-04	2.17 e-04	0.196	-2.33 e-05	2.33 e-04	0.696	-2.16 e-05	5.11 e-05	0.642
Growth (within)	-6.26 e-05	3.10 e-05	0.060	6.54 e-05	1.40 e-04	0.807	3.19 e-05	1.77 e-04	0.707	1.88 e-04	1.60 e-04	0.425	-9.71 e-06	4.26 e-05	0.856
Age	0.072	0.039	0.120	-0.206	0.15	0.214	-0.137	0.192	0.994	-0.229	0.186	0.326	0.049	0.046	0.105
Age^2	-6.62 e-03	3.03 e-03	0.064	0.022	-0.016	0.170	9.21 e-03	0.021	0.653	0.027	0.201	0.270	-4.57 e-03	4.93 e-03	0.104
Mass (between)	4.16 e-05	2.94 e-05	0.131	-7.88 e-05	1.17 e-04	0.332	2.38 e-04	1.48 e-04	0.172	2.63 e-05	1.46 e-04	0.839	8.62 e-06	3.57 e-05	0.999
Mass (within)	**-9.26 e-05**	**3.09 e-05**	**0.001**	**3.34 e-04**	**1.06 e-04**	**0.006**	2.31 e-05	1.33 e-04	0.498	5.00 e-05	1.25 e-04	0.865	2.98 e-05	3.24 e-05	0.222

Estimate, standard error, and *p*-values for growth, age (as a quadratic function), and mass for all docility and activity and exploration in both OF and MIS. We provide both between and within individual estimates for growth and mass.

Figure 4-1



Relationship between personality traits docility (panels a-c), activity (d-i) and exploration (j-o) and growth (difference in grams between mass on the 15th of June and 15th of August), age (in years), and mass at capture (in grams). Activity and exploration were estimated from principal components analyses of behaviours in both open field (OF) and mirror image stimulation (MIS) tests (factors 1 and 2 for OF, factor 1 and 3 for MIS respectively).

Supplementary Methods

Growth: Previous studies show that estimating body mass using best linear unbiased predictors (BLUPs) calculated from a linear mixed effects models are more accurate than using a simple linear regression for individual (Martin & Pelletier, 2011). Using multiple body mass measures per individuals per year, we fitted a linear mixed model with body mass as a function of a quadratic effect of day of the year. We included identity, day, year, and colony as random effects. We extracted both individual intercepts and slopes (BLUPs) to estimate mass on the 1 June and 15 August for each year. Growth is the difference between these two estimates.

Supplementary Table 1.

	Open Field				Mirror Image Stimulation					
Behavior/Traits	1	2	3	4	1	2	3	4	5	6
N heart beats (15 sec.)	0.136	-0.139	**0.601**	-0.519	0.131	-0.18	-0.057	**0.732**	-0.186	-0.258
Defecate	0.021	-0.162	0.404	**0.66**	0.123	-0.036	-0.05	0.041	-0.074	**0.868**
Urinate	0.073	0.128	-0.077	**0.594**	-0.107	-0.023	0.276	-0.086	0.476	0.323
Immediately out	0.008	0.053	**0.713**	0.088	-0.103	0.076	0.054	**0.801**	0.106	0.252
Percent boxes visited	**0.754**	0.34	0.11	0.062	**0.705**	0.263	0.443	0.011	0.092	0.105
N lines crossed	**0.908**	0.199	0.043	0.084	**0.755**	0.299	0.335	-0.044	0.034	0.046
N alarm call	-0.112	-0.071	-0.385	-0.01	0.087	0.031	-0.158	0	**0.841**	-0.183
N jump	**0.615**	-0.384	-0.214	0.152	**0.692**	0.14	-0.177	-0.116	-0.08	0.078
N sniff	0.44	**0.82**	0.042	0.081	0.372	0.208	**0.827**	0.022	-0.023	-0.016
N walk	**0.889**	0.225	0.124	-0.011	**0.822**	0.289	0.314	0.066	-0.034	-0.008
N total look	**0.793**	0.173	0.138	0.029	**0.728**	0.268	0.212	0.045	0.361	-0.017
Proportion sniff	0.335	**0.876**	0.048	0.085	0.264	0.16	**0.874**	-0.01	0.001	-0.017
Proportion walk	**0.914**	0.126	0.158	-0.063	**0.808**	0.187	0.327	0.127	-0.089	-0.026
Proportion look	**-0.849**	-0.43	-0.146	0.015	-0.49	**-0.625**	-0.391	-0.017	0.123	0.048
Latency to approach mirror (sec.)	-	-	-	-	-0.327	**-0.608**	-0.197	-0.066	-0.171	-0.082
Proportion spent at mirror	-	-	-	-	0.14	**0.841**	-0.06	-0.064	0.155	-0.018
N scratch mirror	-	-	-	-	0.426	0.763	0.206	-0.08	-0.097	-0.029
Proportion scratch mirror	-	-	-	-	0.169	**0.877**	0.168	-0.032	-0.11	-0.039
Percent variance explained	42.518	9.896	8.867	7.563	40.351	9.178	6.963	6.555	6.178	5.715
Total variance	42.518	52.413	61.28	68.843	40.351	49.528	56.491	63.047	69.224	74.939

Principle component analysis of behaviours scored in both open field (OF) and mirror image stimulation (MIS) tests. Loadings with variance explained by each component and total overall variance for components with an eigenvalue greater than one. We considered a variable significantly loaded onto a component if it had a value greater than |0.500|. Defecate, urinate, and immediately ran out of the trap were scored as happening with a dichotomous value (0/1).

Supplementary Table 2.

Fixed Effect	Docility	Activity (OF1)	Exploration (OF2)	Activity (MIS1)	Exploration (MIS3)
Intercept	4.15± 0.170***	-0.407± 0.610	-2.146± 0.8469*	0.178± 0.8016	1.752± 0.1799
Date	-9.76E-04± 6.42E-04	4.33E-03± 2.65E-03	8.25E-03± 3.56E-03*	1.52E-03± 3.37E-03	9.60E-04± 7.87E-04
Days between trials	2.30E-03± 1.04E-03*	-1.61E-02± 4.83E-03***	-6.28E-03± 5.90E-03	-1.14E-02± 5.23E-03**	2.22E-04± 1.47E-03
Predator Presence	1.11E-02± 2.50E-02	3.711± 0.716***	0.5283± 0.9534	2.131± 0.976*	7.20E-02± 0.214
Time (PM)	-0.146± 2.56E-02***	3.83E-02± 9.80E-02	0.1386± 0.1193	0.1089± 0.1085	-3.67E-03± 3.04E-02
Sex (M)	2.88E-03± 3.76E-02	-0.132± 0.1089	-7.71E-02± 0.1408	-4.52E-02± 0.1504	-8.95E-03± 3.32E-02
Pedestrian Presence	-2.63E-03± 1.64E-03*	-9.24E-03± 5.56E-03	1.10E-02± 7.60E-03*	-2.32E-02± 8.09E-03*	1.47E-03± 1.66E-03
Trial	3.13E-02± 4.52E-03***	-1.56E-01± 4.82E-02***	-1.34E-01± 5.87E-02	-2.40E-02± 5.27E-02	-6.62E-03± 1.48E-02

<0.05; ** < 0.01; *** <0.001

Coefficients and standard errors for fixed effects of models for docility and activity and exploration in both open field (OF) and mirror image stimulation (MIS) tests with growth, age, and mass.

References

Armitage, K.B. & Downhower, J.F. 1974. Demography of yellow-bellied marmot populations. *Ecology* **55**: 1233–1245.

Baayen, R.H., Davidson, D.J. & Bates, D.M. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* **59**: 390–412.

Bates, D., Maechler, M. & Bolker, B. 2012. *lme4: Linear mixed-effects models using S4 classes (2011). R package version 0.999375-42*.

Bergmüller, R. & Taborsky, M. 2010. Animal personality due to social niche specialisation. *Trends Ecol. Evol.* **25**: 504–511.

Biro, P.A., Beckmann, C. & Stamps, J.A. 2010. Small within-day increases in temperature affects boldness and alters personality in coral reef fish. *Proc. R. Soc. B Biol. Sci.* **277**: 71–77.

Biro, P.A. & Post, J.R. 2008. Rapid depletion of genotypes with fast growth and bold personality traits from harvested fish populations. *Proc. Natl. Acad. Sci.* **105**: 2919–2922.

Biro, P.A. & Stamps, J.A. 2008. Are animal personality traits linked to life-history productivity? *Trends Ecol. Evol.* **23**: 361–368.

Blumstein, D.T. & Daniel, J.C. 2007. *Quantifying behavior the JWatcher way*. Sinauer Associates Inc.

Blumstein, D.T., Ozgul, A., Yovovich, V., Van Vuren, D.H. & Armitage, K.B. 2006. Effect of predation risk on the presence and persistence of yellow-bellied marmot (*Marmota flaviventris*) colonies. *J. Zool.* **270**: 132–138.

Blumstein, D.T., Wey, T.W. & Tang, K. 2009. A test of the social cohesion hypothesis: interactive female marmots remain at home. *Proc. R. Soc. B Biol. Sci.* **276**: 3007–3012.

Bouwhuis, S., Quinn, J.L., Sheldon, B.C. & Verhulst, S. 2014. Personality and basal metabolic rate in a wild bird population. *Oikos* **123**: 56–62.

Careau, V., Bininda-Emonds, O.R.P., Thomas, D.W., Réale, D. & Humphries, M.M. 2009. Exploration strategies map along fast–slow metabolic and life-history continua in muroid rodents. *Funct. Ecol.* **23**: 150–156.

Careau, V., Thomas, D., Humphries, M.M. & Réale, D. 2008. Energy metabolism and animal personality. *Oikos* **117**: 641–653.

Clark, C.W. 1994. Antipredator behavior and the asset-protection principle. *Behav. Ecol.* **5**: 159–170.

Clobert, J., Oppliger, A., Sorci, G., Ernande, B., Swallow, J.G. & Garland, T. 2000. Trade-offs in phenotypic traits: endurance at birth, growth, survival, predation and susceptibility to parasitism in a lizard, *Lacerta vivipara*. *Funct. Ecol.* **14**: 675–684.

Dammhahn, M. 2012. Are personality differences in a small iteroparous mammal maintained by a life-history trade-off? *Proc. R. Soc. B Biol. Sci.* **279**: 2645–2651.

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

Dingemanse, N.J. & Wolf, M. 2010. Recent models for adaptive personality differences: a review. *Philos. Trans. R. Soc. B Biol. Sci.* **365**: 3947–3958.

Duchesne, M., Côté, S.D. & Barrette, C. 2000. Responses of woodland caribou to winter ecotourism in the Charlevoix Biosphere Reserve, Canada. *Biol. Conserv.* **96**: 311–317.

Edenbrow, M. & Croft, D.P. 2011. Behavioural types and life history strategies during ontogeny in the mangrove killifish, *Kryptolebias marmoratus*. *Anim. Behav.* **82**: 731–741.

Frase, B.A. & Hoffmann, R.S. 1980. Marmota flaviventris. *Mamm. Species* 1–8.

Gyuris, E., Feró, O., Tartally, A. & Barta, Z. 2011. Individual behaviour in firebugs (*Pyrrhocoris apterus*). *Proc. R. Soc. B Biol. Sci.* **278**: 628–633.

Huang, B., Wey, T.W. & Blumstein, D.T. 2011. Correlates and consequences of dominance in a social rodent. *Ethology* **117**: 573–585.

Krause, E.T. & Naguib, M. 2011. Compensatory growth affects exploratory behaviour in zebra finches, *Taeniopygia guttata*. *Anim. Behav.* **81**: 1295–1300.

Lacy, K.E. & Martins, E.P. 2003. The effect of anthropogenic habitat usage on the social behaviour of a vulnerable species, *Cyclura nubila*. *Anim. Conserv.* **6**: 3–9.

Laskowski, K.L. & Pruitt, J.N. 2014. Evidence of social niche construction: persistent and repeated social interactions generate stronger personalities in a social spider. *Proc. R. Soc. B Biol. Sci.* **281**.

Li, C., Monclús, R., Maul, T.L., Jiang, Z. & Blumstein, D.T. 2011. Quantifying human disturbance on antipredator behavior and flush initiation distance in yellow-bellied marmots. *Appl. Anim. Behav. Sci.* **129**: 146–152.

Luttbeg, B. & Sih, A. 2010. Risk, resources and state-dependent adaptive behavioural syndromes. *Philos. Trans. R. Soc. B Biol. Sci.* **365**: 3977–3990.

Martin, J.G.A. & Pelletier, F. 2011. Measuring growth patterns in the field: effects of sampling regime and methods on standardized estimates. *Can. J. Zool.* **89**: 529–537.

Martin, J.G.A. & Réale, D. 2008. Temperament, risk assessment and habituation to novelty in eastern chipmunks, *Tamius striatus*. *Anim. Behav.* **75**: 309–318.

Nicolaus, M., Tinbergen, J.M., Bouwman, K.M., Michler, S.P., Ubels, R., Both, C., *et al.* 2012. Experimental evidence for adaptive personalities in a wild passerine bird. *Proc. R. Soc. B Biol. Sci.* **279**: 4885–4892.

Petelle, M.B., McCoy, D.E., Alejandro, V., Martin, J.G. & Blumstein, D.T. 2013. Development of boldness and docility in yellow-bellied marmots. *Anim. Behav.* **86**: 1147–1154.

Pianka, E.R. & Parker, W.S. 1975. Age-specific reproductive tactics. *Am. Nat.* 453–464.

Pinheiro, J.C. & Bates, D.M. 2000. Mixed-effects models in S and S-PLUS Springer. *N.Y.*

Pottinger, T.G. 2006. Context dependent differences in growth of two rainbow trout (*Oncorhynchus mykiss*) lines selected for divergent stress responsiveness. *Aquaculture* **256**: 140–147.

Pruitt, J.N. & Ferrari, M.C.O. 2011. Intraspecific trait variants determine the nature of interspecific interactions in a habitat-forming species. *Ecology* **92**: 1902–1908.

Rankin, C.H., Abrams, T., Barry, R.J., Bhatnagar, S., Clayton, D.F., Colombo, J., *et al.* 2009. Habituation revisited: An updated and revised description of the behavioral characteristics of habituation. *Neurobiol. Learn. Mem.* **92**: 135–138.

RDevelopment, C. 2012. TEAM 2009: R: A Language and Environment for Statistical Computing. *Vienna Austria Internet Httpwww R-Proj. Org*.

Réale, D., Gallant, B.Y., Leblanc, M. & Festa-Bianchet, M. 2000. Consistency of temperament in bighorn ewes and correlates with behaviour and life history. *Anim. Behav.* **60**: 589–597.

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

Réale, D., Martin, J., Coltman, D.W., Poissant, J. & Festa-Bianchet, M. 2009. Male personality, life-history strategies and reproductive success in a promiscuous mammal. *J. Evol. Biol.* **22**: 1599–1607.

Réale, D., Reader, S.M., Sol, D., McDougall, P.T. & Dingemanse, N.J. 2007. Integrating animal temperament within ecology and evolution. *Biol. Rev.* **82**: 291–318.

Sih, A., Bell, A.M., Johnson, J.C. & Ziemba, R.E. 2004. Behavioral syndromes: an integrative overview. *Q. Rev. Biol.* **79**: 241–277.

Sinervo, B. & Calsbeek, R. 2010. Behavioral concepts of selection. *Evol. Behav. Ecol. Oxf. Univ. Press N. Y.* 32–45.

Smith, B.R. & Blumstein, D.T. 2008. Fitness consequences of personality: a meta-analysis. *Behav. Ecol.* **19**: 448–455.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & Van Der Linde, A. 2002. Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64**: 583–639.

Stamps, J.A. 2007. Growth-mortality tradeoffs and "personality traits" in animals. *Ecol. Lett.* **10**: 355–363.

Sundström, L.F., Petersson, E., Höjesjö, J., Johnsson, J.I. & Järvi, T. 2004. Hatchery selection promotes boldness in newly hatched brown trout (*Salmo trutta*): implications for dominance. *Behav. Ecol.* **15**: 192–198.

Svendsen, G.E. & Armitage, K.B. 1973. Mirror-image stimulation applied to field behavioral studies. *Ecology* **54**: 623.

Vainikka, A., Rantala, M.J., Niemelä, P., Hirvonen, H. & Kortet, R. 2011. Boldness as a consistent personality trait in the noble crayfish, *Astacus astacus*. *Acta Ethologica* **14**: 17–25.

Ward, A.J.W., Thomas, P., Hart, P.J.B. & Krause, J. 2004. Correlates of boldness in three-spined sticklebacks (*Gasterosteus aculeatus*). *Behav. Ecol. Sociobiol.* **55**: 561–568.

Wolf, M., van Doorn, G.S., Leimar, O. & Weissing, F.J. 2007. Life-history trade-offs favour the evolution of animal personalities. *Nature* **447**: 581–584.

Wolf, M. & Weissing, F.J. 2010. An explanatory framework for adaptive personality differences. *Philos. Trans. R. Soc. B Biol. Sci.* **365**: 3959–3968.

CHAPTER 5

HERITABILITY AND GENETIC CORRELATIONS OF PERSONALITY TRAITS IN YELLOW-BELLIED MARMOTS

Abstract

Describing and quantifying animal personality is now an integral part of behavioral studies because individually distinctive traits may have ecological and evolutionary consequences. Yet, to fully understand how personality traits may respond to selection, one must understand the underlying heritability and genetic correlations between traits. Most studies that have investigated the additive genetic variance of personality traits typically find a moderate amount of heritable variation, but few studies have been conducted on wild populations. Estimating heritability in the wild is important because environmental conditions reveal the often reduced, additive genetic variance found in nature. In addition, to understand how a population may respond to selection, one must identify possible constraints caused by genetic correlations. We estimated the additive genetic variance of docility, exploration, and activity in a wild population of yellow-bellied marmots (*Marmota flaviventris*), and the additive genetic variance for exploration and activity in a second context. We found little to no significant heritability in these traits (0.033-0.151). We found phenotypic correlations were explained by both genetic and permanent environment correlations but not correlations between maternal effects. This is one of a handful of studies to take a quantitative genetic approach to understand personality traits in the wild, and thus, gives insights into the evolution and maintenance of personality.

Introduction

Individuals from many different taxa have been shown to behave in consistent, individually different ways (Gosling, 2001) —a phenomenon referred to as personality— which may have important ecological and evolutionary consequences (Réale *et al.*, 2007). For example, variation in personality traits is often associated with alternative strategies in life history (Réale *et al.*, 2000; Wolf *et al.*, 2007), and can have important effects on fitness (Smith & Blumstein, 2008). Thus, these consistent differences may have long-term demographic consequences for populations. Furthermore, the substantial phenotypic variation within personality traits suggests that variation is actively maintained within populations (Boon *et al.*, 2007). To understand the maintenance of this variation, it is important to understand the additive genetic variation upon which selection may act. Additionally, many personality traits are phenotypically correlated with each other and create what are referred to as behavioral syndromes (Sih *et al.*, 2004). Such syndromes may constrain selection and prevent the erosion of genetic variation under constant selection. Therefore, to understand the potential response to selection of a trait within a population, one must know the heritability of that trait, as well as the potential genetic constraints generated by genetic correlations.

Few studies have investigated the genetic and environmental sources of co(variances) of behavior and personality (Stirling *et al.*, 2002) despite the importance of these effects on evolution (Falconer & Mackay, 1996; Lynch & Walsh, 1998). Quantitative genetics is based on the theory that complex traits are based on not just a small number of genes, but many genes. These genes make up the additive genetic variance of a trait. As additive genetic variance increases compared to phenotypic variance, the heritability, or h^2, also increases. Heritability is part of the breeder's equation, and is important in estimating how a trait will react to selection (Falconer & Mackay, 1996). Few studies have estimated the heritability of personality traits in the wild, and thus, we know very little about how personality reacts to selection in nature. Indeed, lab-based estimates seemingly over-estimate additive genetic variation when compared to the low to moderate heritabilities reported in the wild (van Oers *et al.*, 2005; Sinn *et al.*, 2006; Lea *et al.*, 2010; Taylor *et al.*, 2012) suggesting that the environment has a large effect on phenotypes. In one of the best examples of heritability of personality in a wild population, (Taylor *et al.*, 2012), found that a population of red squirrels (*Tamiasciurus hudsonicus*) had relatively small heritabilities (0.08-0.12) in a number of commonly measured personality traits. They also found relatively small maternal and permanent environment effects, but found larger genetic and maternal genetic correlations. Two studies by Réale et al. (2000, 2009) found moderate to high heritability in big horn sheep (*Ovis canadensis*). However, they detected no maternal effects in their population.

In addition to quantifying the heritability of a trait, one must understand what constrains trait evolution. Traits are often not independent of one another due to linkage disequilibrium or pleiotropic effects. Selection is therefore multivariate, influencing multiple traits at once (Lande & Arnold, 1983). Only by understanding the genetic architecture and the underlying correlations can we understand selection and the evolution of traits. These correlations between personality traits, or behavioral syndromes, may have an underlying genetic cause (Dochtermann, 2010) and may explain the suboptimal behavior of some individuals (Sih *et al.*, 2004). Thus, to understand how personality traits may respond to selection and evolve, we must understand the magnitude and directions of genetic correlations as well.

We quantified the additive genetic, maternal, and permanent environment variances and the correlations of four personality traits – docility, activity, sociability, and exploration – in a wild population of yellow-bellied marmots (*Marmota flaviventris*- hereafter referred to as marmots). Behavioral syndromes are hypothesized to have an underlying genetic component, but this assumption is often not tested. Other potential mechanisms include correlations due to permanent environment or maternal effects. Furthermore, no studies to our knowledge have tested whether personality traits measured under different contexts share the same underlying genetic causes. Like other studies in the wild (Taylor *et al.*, 2012), we expect heritability to be relatively small because of high environmental variation. Although we expect phenotypic correlations among and within personality traits, we have no *a priori* hypotheses about the underlying architecture of those correlations.

Methods

Study species and sites

Yellow-bellied marmots are large (3-5kg), semifossorial, sciurid rodents, native to North America, that live in colonies that consist of one or more matrilineal groups (Frase & Hoffmann, 1980). Marmots are active from mid-April to mid-October and hibernate through the winter (Blumstein *et al.*, 2006). We differentiate three age categories: juveniles, which are young of the year; yearlings, individuals that have survived their first winter; and adults, individuals that have survived their second winter and are reproductively mature. Our study population is located in the upper East River Valley, Gunnison, Colorado, the site of Rocky Mountain Biological Laboratory (RMBL). We regularly trap multiple colonies in and around the RMBL, which is located in Gothic, Colorado (38° 57' 29" N; 106° 59' 06" W). This population has been followed since 1962 (Armitage, 2010; Blumstein, 2013), and the individual behavior used in this study having been collected since 2002 (Petelle *et al.*, 2013).

Pedigree

We assigned parentage using DNA collected from individuals studied from 2002-2012. Detailed methods are described in (Olson & Blumstein, 2010). Briefly, however, we extracted DNA using Qiagen QIAamp DNA MINI kits and genotyped individuals at 12 microsatellites. Alleles were visualized and scored using GENEMAPPER, and parentage was assigned using CERVUS 3.0 (Kalinowski *et al.*, 2007). Juveniles were trapped the first time they emerged out of the maternal burrow. This allowed us to behaviorally match juveniles to mothers. To confirm behavioral assignment, we ran CERVUS to match maternity and paternity to juveniles using a maximum likelihood method at 95% trio confidence. Since 2002, we have genotyped 1432 individuals from 136 dams and 71 sires.

Quantifying Personality

Docility is a commonly measured personality trait, and is a measure of how an individual reacts to being trapped and handled (Réale *et al.*, 2000). We quantified docility in 920 individuals with data collected during 7904 trapping events from 2002 through 2012. At each trapping event we dichotomously (0/1) scored whether individuals struggled in the trap, tooth chattered, alarm called, struggled in the bag, and hesitated to walk into the handling bag. These were summed and subtracted from the total potential score. Thus, an individual who scored a 5 is considered docile during that trapping event while an individual who scores a 0 is non-docile.

During the 2010-2012 active seasons, we tested 183 individuals in 508 open-field and mirror image stimulation test. After individuals had been regularly trap processed (weighed, left hind foot measured, sexed, ear tags checked and replaced if required, feces collected if present), they were brought to a shaded arena for testing. Open-field (MIS) and mirror image stimulation (MIS) tests were conducted in an opaque arena measuring 91.4 cm^3 made of 0.47 cm opaque PVC sheeting with a wire mesh top to prevent escape. A mirror (30.5 x 61.0 cm) was placed at the base of one side of the arena and covered with an opaque sliding door. A door (61.0 cm^2) was cut out of the opposite side. Sixteen equal squares were drawn on the floor of the arena to quantify activity and exploration. Individuals were gently placed in the middle of the arena and their subsequent behavior was video recorded. The first three minutes were the OF test where the marmots were able to explore the arena without obstruction. After the OF test, the mirror was exposed and the following three minutes constituted the MIS test. Marmots were gently coaxed back into a trap and transported and released where they were originally trapped.

OF and MIS behavior was scored using the event recorder JWatcher (Blumstein & Daniel, 2007), which allowed us to quantify the duration and frequency of the following behaviors; walk (quadrapedal and bipedal), look (quadrapedal and bipedal), jump, alarm call, and sniffing/smelling. For MIS only, we also included scratching/pawing at the mirror. We also quantified the number of squares each individual entered and the proportion of squares entered (See Petelle & Blumstein, 2014 for full methods).

Statistical Analysis

Rather than analyzing each OF/MIS behavior separately, we chose to reduce the number of correlated traits using a principal components analysis with varimax rotation. OF and MIS were analyzed separately. Components with eigenvalues greater than 1.0 were retained for further analysis.

We estimated additive genetic, permanent environment, maternal, and year effects for the resulting OF and MIS components and docility using an animal model with a Bayesian approach (Wilson *et al.*, 2010). We fitted the generalized linear mixed effects models using a Markov Chain Monte Carlo. We first fit each personality trait with a univariate model with the fixed effects of sex and age class and other fixed effects that had previously been shown to significantly influence personality traits (Petelle et al. in review). The sociability component had previously not been analyzed, so we included sex, age class, Julian date, pedestrian traffic, predator pressure, days between trial, and trial number as fixed effects (see Petelle et al., 2013 for methods on the calculation of pedestrian and predator pressure). We included individual twice, dam, and year as random effects. In this case, individual is included twice in order to separate the additive genetic and permanent environmental variances. One individual term is connected to the pedigree and estimates the additive genetic variance while the other individual term estimates the permanent environment. Variance parameters were estimated as the posterior mode with 95% credible intervals based on the posterior distribution of the parameter. Posterior distribution of heritability was estimated with the equation h^2 = V_A/V_P. In this equation, h^2 is heritability, V_A is the additive genetic variance, and V_P is the total phenotypic variance. Other ratios were calculated the same way except for repeatability, which is the sum of both additive and permanent environmental effects (repeatability = $(V_A + V_{PE})/V_P$). Since variance parameters are bounded above zero, we estimated importance of random effects by looking at the deviance information criteria (DIC) (Spiegelhalter *et al.*, 2002). DIC is analogous to the Bayesian version of Akaike information criterion (AIC). For this reason, we used a delta DIC value of 2 to identify important random effects. To do so, we removed random effects one at a time from the full model and estimated the DIC.

To estimate genetic correlations, we fit bivariate models for each pair of personality traits. We rescaled the covariance into correlations using the equation ($r = Cov_{a,b} / \sqrt{V_a V_b}$).

For all analyses, we used the package MCMCglmm (Hadfield, 2010) in R (RDevelopment, 2012). For univariate models, the posterior distribution was sampled every 500 iterations with a burning of 30,000 for a total of 530,000 samples. The bivariate models were sampled every 1000 iterations with a burning of 30,000 for a sample of 1,030,000. Mixing of chain was assessed visually and the autocorrelation was below 0.05 for all parameters. We used non-informative inverse-wishart for all models.

Results

We extracted four and six principle components for OF and MIS tests, respectively. After reviewing the component loadings, we identified two OF and three MIS components that corresponded with personality traits. The first component in each test was labeled activity, the second and third component of the OF and MIS test, respectively, were identified as exploratory, and the second MIS component was labeled sociability (Supplementary Table 1 for component loadings). The two OF components explained approximately 52% of the variance while the three MIS components explained 56% (Petelle *et al.*, 2013).

The additive genetic variance for docility was relatively small, but nonetheless significant (h^2 = 0.07; 95% CI = 0.031 - 0.128; ΔDIC = 3.930). We also found activity (OF1) (0.151; 0.026 – 0.258; 11.676) and exploration (MIS3) (0.067; 0.013 – 0.137; 4.529) to be heritable. All estimates of heritability are given in Table 1. While the delta DIC value for sociability (MIS2) (0.039; 0.013 – 0.181; 1.943) was not above our threshold for inclusion, we include it in our discussion as a heritable trait (See Supplementary Table 2 for all ΔDIC).

The variation attributed to the permanent environment was also small in most cases (Table 1), but was larger than the additive genetic variance in docility, exploration (OF2), activity (MIS1), sociability (MIS2), and exploration (MIS3) (Table 1). Maternal effects were also small, but docility (me^2 = 0.041; 95%CI = 0.020 – 0.069; ΔDIC = 11.31) and sociability (0.058; 0.013 – 0.153; 4.015) had large ΔDIC . Year effects similarly were small and significant for docility (ye^2 = 0.024; 95% CI = 0.0119 – 0.070; ΔDIC = 58.63), and both activity (ye^2_{OF} = 0.088; 95% CI = 0.013 – 0.153; ΔDIC =

16.495; ye^2_{MIS} = 0.077; 0.143 – 0.514; 20.475) and exploration (ye^2_{OF} = 0.147; 95% CI = 0.015 – 0.209; ΔDIC = 26.051; ye^2_{MIS} = 0.062; 0.019 – 0.195; 11.298) in each context (Table 1).

We found a number of behavioral syndromes, or phenotypic correlations, among personality traits. As expected, we found a positive phenotypic correlation between activity in the OF and MIS tests (r_P = 0.571; 95% CI = 0.303 – 0.741). We also found a positive correlation between both OF and MIS activity and sociability (OF/sociability: r_P = 0.450; 95% CI = 0.206 – 0.712 and MIS/sociability: r_P = 0.483; 0.185 – 0.687). Finally, we found one negative correlation between activity in the OF test and docility (r_P = -0.301; 95%CI = -0.571 to -0.074) (for full P-matrix see Supplementary Table 3).

We then investigated the potential genetic, permanent environment, and maternal correlations that may be the underlying cause of these phenotypic correlations. We found only one significant genetic correlation – activity in the OF test and sociability (r_G = 0.673; 95% CI = 0.005 – 0.833) (for full G-matrix see Supplementary Table 4). We did, however, find a number of permanent environmental correlations between activity in the OF and MIS tests (r_{PE} = 0.641; 95% CI = 0.095-0.862), and, interestingly, between docility and exploration in the MIS test (0.521; 0.070 – 0.806) (Supplementary Table 5). We found no maternal correlations between or within traits (Supplementary Table 6). Effects with 95% CI excluding zero were deemed significant. It should be noted that there are a number of phenotypic, genetic, and permanent environment correlations that are moderate to high but were not significant because of large 95% confidence intervals and almost excluded zero.

Discussion

The maintenance of personality variation is an important question in behavioral ecology and evolution. This is because personality traits may be linked to life history syndromes (Wolf et al., 2007; Biro & Stamps, 2008), and because they may have both fitness consequences (Smith & Blumstein, 2008), and influence population demography (Armitage, 1986). Personality variation may be maintained because there are multiple optima on a fitness landscape, or because there is fluctuating selection over time or space (Boon *et al.*, 2007). However, in each of these scenarios, personality must be heritable to evolve.

Our results suggest a number of conclusions. First, some of the personality traits we investigated have low heritability, and therefore variation has some underlying genetic origin. Second, there are a number of distinct phenotypic correlations. Activity is correlated across context, activity in both contexts is correlated with sociability, and activity is negatively associated with docility. Finally, potential evolutionary change for some of these traits is likely to be constrained by underlying genetic or permanent environment correlations. Because some of these correlations are due to the shared permanent environment, this suggests that natural selection has coupled these traits together, and therefore syndromes at the phenotypic level may be an adaptive strategy for this population.

We found none to low heritability in all of our personality traits. Although these estimates are relatively small when compared with other behaviors (Stirling *et al.*, 2002), heritability estimates in personality traits vary widely, but estimates are generally smaller when estimated in the wild than in captivity (Sinn *et al.*, 2006; Taylor *et al.*, 2012). A study on wild dumpling squid (*Euprymna tasmanica*) tested in the laboratory found non-significant heritability of behavioral traits in a number of different foraging contexts, but did identify significant heritability (h^2 = 0.2-0.8) in anti-predator behaviors (Sinn *et al.*, 2006). And, like our study, Taylor et al. (2012) found low heritability in docility (h^2 = 0.09), aggression (h^2 = 0.12), and activity (h^2 = 0.08) in wild red squirrels. Thus, personality traits measured in the wild, despite having moderate repeatabilities, may generally have low heritabilities. Low heritability might suggest that these traits are linked to fitness and genetic variation has been eroded (Falconer & Mackay, 1996; Kruuk *et al.*, 2000). An equally likely hypothesis is those residual and phenotypic variances covary. This is because when phenotypic variance increases, the residual variance also increases. This covariation results in the additive genetic component of variance explaining less of the total phenotypic variance, a process that reduces heritability estimates (Stirling *et al.*, 2002). Repeatability sets the upper limit to heritability (Falconer & Mackay, 1996) and our results show that repeatability in our population is often much higher than heritability. Our estimates of repeatability were low to moderate, and are generally comparable to most repeatability estimates of personality in the wild, but are lower than most estimates of behavioral repeatability (Bell *et al.*, 2009).

We also found that permanent environment effects were, in some cases, larger than heritability estimates. This suggests that, for those traits where this pattern is found, the shared environment potentially plays a much larger role in accounting for personality variation than the underlying genes. This is not surprising because, in the wild, the environment often has a large effect on trait variation. Indeed, other studies of personality in the wild reported substantial and significant permanent environmental effects (Réale *et* *al.*, 2009; Taylor *et al.*, 2012). We also found small maternal effects in docility, and found a small, but not significant maternal effect on sociability (MIS2). Maternal effects can have long-term consequences on individuals (Reinhold, 2002; Weaver *et al.*, 2004; Räsänen & Kruuk, 2007). However, few studies of personality have estimated these effects in wild populations (Réale et al., 2009; Taylor et al., 2012). These maternal effects on sociability can have long lasting effects on this population. Social, or more well-connected, female marmots are more likely to remain in their natal colony (Blumstein *et al.*, 2009), and the recruitment of individuals into a colony can have large demographic consequences (Armitage, 1986). Thus, maternal effects on sociability may have wide-ranging effects on this population of marmots. Like our study, Taylor et al. (2012) found low heritability in their population while Réale et al. (2009) found none.

Estimating heritability and different environmental and maternal effects alone is insufficient to understand how traits may respond to selection. How those traits covary at the phenotypic and genetic level is equally important to understanding evolutionary potential. We found a number of traits correlated at the phenotypic level. Activity was correlated across contexts, as it should, if tests are measuring the same trait. Activity was also positively associated with sociability. This correlation could arise if more active individuals are coming into contact with more individuals, or because there is a high degree of betweenness between different social groups (Krause *et al.*, 2010). Betweenness is a social network metric that measures the centrality of an individual based on the shortest paths between pairs of individuals in that group. Thus, if an individual connects two groups and has connections within each group, they have a high level of betweenness (Wey *et al*. 2008). The exact ecological consequences of these traits are unknown at this time, and future work should focus on how sociability in the MIS is associated with sociability in a natural setting using social network metrics. Finally, there was a negative correlation between activity in the OF test and docility. This also makes intuitive sense as docility is composed of measures of activity in trapping and handling (struggling in trap or handling). Thus more active individual in the OF test may also be more active in the trap and therefore receive lower docility scores.

We found only one genetic correlation underlying these phenotypic correlations. The genetic correlation between activity in the OF test and sociability in the MIS test means that these traits may constrain the independent evolution of each other. We are unable to determine whether these correlations are from pleiotropy or linkage disequilibrium.

We also detected two permanent environment correlations. This suggests that the shared environment, either at the colony or site level, influences the formation of syndromes. The positive permanent environment between activity between OF and MIS makes intuitive sense methodologically and because a common environment may select for a context-general activity syndrome. However, because these traits are not correlated at the genetic level, they are not considered the same trait. But since they are phenotypically correlated at the individual level, there must be something during ontogeny that influences activity in both contexts. We are unsure of what this might be, but some aspect of the shared environment during development profoundly influenced our activity measurements. The positive permanent environment correlation between docility and exploration in the MIS tests is interesting, and this is the first study, to our knowledge, to detect such a correlation.

Although many of the phenotypically correlated traits did not have underlying genetic correlations, we did find moderate genetic and permanent environment correlations that were non-significant. These correlations had large 95% confidence intervals and were close to excluding zero. This suggests two possible explanations; that these traits are in fact correlated at the genetic or permanent environmental levels and we do not have enough power to estimate correlations or reduce confidence intervals, or these traits are in fact not correlated at specific levels. This implies that traits are truly independent of one another as well as traits measured across contexts are not the same trait despite the fact that we are measuring the same expressed behaviors (our PCA components labeled exploration – OF2 and MIS3). This is important, because most studies assume that they are measuring the same trait between contexts (Carter *et al.*, 2013), but these traits need not be genetically correlated. More generally, this highlights an issue: researchers must be wary that the personality traits they measure in different contexts may not have any underlying genetic correlation and thus selection may act independently on these traits. Additionally, tests that are done outside of the individual's natural setting (here trapping behaviors and OF and MIS tests) should be done with the realization that these tests may constrain the expressed behavior of animals and that we may not be measuring the true personality trait.

How these correlations, or syndromes, are formed is an active area of evolutionary research (Dochtermann, 2010). Two hypotheses, constraint or adaptation, are used to describe the presence of these syndromes. The constraint hypothesis states that personality traits have an underlying genetic or physiological cause (Sih *et al.*, 2004), and that this correlation prevents traits from reaching their own independent optima

(Dochtermann, 2010). The adaptive hypothesis states that natural selection forms these suites of behaviors to be adaptive in that population's specific environment (Bell, 2005; Dingemanse *et al.*, 2007).

Few studies have tested for genetic correlations between personality traits in wild populations (Bell, 2005; Réale *et al.*, 2009; Taylor *et al.*, 2012), and all of these studies found genetic correlations between traits. Interestingly, Bell (2005) found that correlation strength changed between populations of three-spined sticklebacks (*Gasterosteus aculeatus*) with or without strong predation pressure. This result suggests that correlations between traits might be an adaptive strategy (Wilson, 1998). Indeed, she found a fitness consequence to the correlation between traits that is linked to life history strategies, a finding that suggests that the link is adaptive.

The generally small genetic effects that we report and that are reported in other studies, illustrates the large influence of the environment on trait variation. Indeed, the magnitude of environmentally-caused variation means that large sample sizes are needed to estimate genotypic/phenotypic correlations in wild populations (Kruuk, 2004). While our trap-related sample sizes were very large (>7000 trapping events), we conducted substantially fewer OF and MIS experiments and this provides a constraint on estimating effects. Nonetheless, with those somewhat smaller samples sizes we were able to estimate other effects in our mixed models, which further highlights the relatively small amount of genetic variation in these traits. Environment, in our population, has a predominantly large influence on personality. Furthermore, our results suggest that there are few underlying genetic correlations between some traits, and those permanent environment correlations, or a shared environment, can also have an effect on phenotypic correlations.

This provides evidence in support of the adaptive hypothesis that the environment can generate these correlations. Studies at just the phenotypic level also suggest an adaptive strategy because differences in correlations were found in 12 populations of three-spined sticklebacks where predation differed between populations (Dingemanse *et al.*, 2007).

Thus, the lack of syndromes in a population suggests that a correlation between traits would potentially be non-adaptive. Moreover, the lack of any syndrome also allows these traits to evolve independently of one another and reach potential independent optima (Roff & Fairbairn, 2007). Because only one population was used, we are unable to understand the potential mechanisms behind the correlation or independence of these traits, but future studies of this population will evaluate the fitness consequences and selection to understand how variation is maintained in the population.

Acknowledgements

We would like to thank all the marmoteers that helped collect data over the years, and specifically we would like to thank Amanda Lea for help with the pedigree and Leon Chong for help in the field. M.B.P. was funded by two U.S. Department of Education GAANN Fellowships, an NSF GK-12 Fellowship, and UCLA. D.T.B was supported by the National Geographic Society, UCLA (Faculty Senate and the Division of Life Sciences), a Rocky Mountain Biological Laboratory research fellowship, and by the NSF (IDBR-0754247 and DEB-1119660 to D.T.B., as well as DBI 0242960 and 0731346 to the Rocky Mountain Biological Laboratory).

Table 5-1.

Heritability ($h^2 = V_A/V_P$), permanent environment effects ($PE = V_{PE}/V_P$), maternal effects ($m^2 = V_{ME}/V_P$), year effects ($YE = V_{YE}/V_P$), residual effects, and repeatability ($V_A + V_{PE}/V_P$) for docility, activity and exploration in both contexts, and sociability. All effects are given with the equivalent of 95% confidence intervals in parentheses.

Trait	h^2	PE	m^2	YE	Residual	Repeatability
Docility	0.070 (0.031 - 0.128)	0.111 (0.070 - 0.145)	0.041 (0.020 - 0.069)	0.024 (0.012 - 0.070)	0.744 (0.691 - 0.778)	0.180 (0.151 - 0.222)
Activity (OF1)	0.151 (0.026 - 0.258)	0.063 (0.013 - 0.153)	0.034 (0.004 - 0.111)	0.088 (0.013 - 0.153)	0.594 (0.276 - 0.753)	0.198 (0.067 - 0.340)
Exploration (OF2)	0.033 (0.004 - 0.131)	0.071 (0.015 - 0.209)	0.052 (0.011 - 0.142)	0.147 (0.015 - 0.209)	0.594 (0.238 - 0.754)	0.103 (0.039 - 0.286)
Activity (MIS1)	0.037 (0.007 - 0.228)	0.300 (0.143 - 0.514)	0.033 (0.007 - 0.096)	0.077 (0.143 - 0.514)	0.413 (0.213 - 0.551)	0.412 (0.190 - 0.544)
Sociability (MIS2)	0.039 (0.013 - 0.137)	0.051 (0.013 - 0.116)	0.058 (0.013 - 0.153)	0.058 (0.013 - 0.116)	0.709 (0.436 - 0.823)	0.125 (0.048 - 0.216)
Exploration (MIS3)	0.067 (0.012 - 0.181)	0.070 (0.019 - 0.195)	0.037 (0.009 - 0.107)	0.062 (0.019 - 0.195)	0.628 (0.346 - 0.806)	0.165 (0.063 - 0.311)

Supplementary Table 1.

Principle component analysis of open field (OF) and mirror image stimulation (MIS) tests. Components were Varimax rotated. Variance for each component and total variance explained are at the bottom of the table. We considered any variable over |0.500| as being significantly loaded onto that component. Significant loadings are bolded.

	Open Field				Mirror Image Stimulation					
Behavior/Traits	1	2	3	4	1	2	3	4	5	6
N heart beats (15 sec.)	0.136	-0.139	**0.601**	-0.519	0.131	-0.18	-0.057	**0.732**	-0.186	-0.258
Defecate	0.021	-0.162	0.404	**0.66**	0.123	-0.036	-0.05	0.041	-0.074	**0.868**
Urinate	0.073	0.128	-0.077	**0.594**	-0.107	-0.023	0.276	-0.086	0.476	0.323
Immediately out	0.008	0.053	**0.713**	0.088	-0.103	0.076	0.054	**0.801**	0.106	0.252
Percent boxes visited	**0.754**	0.34	0.11	0.062	**0.705**	0.263	0.443	0.011	0.092	0.105
N lines crossed	**0.908**	0.199	0.043	0.084	**0.755**	0.299	0.335	-0.044	0.034	0.046
N alarm call	-0.112	-0.071	-0.385	-0.01	0.087	0.031	-0.158	0	**0.841**	-0.183
N jump	**0.615**	-0.384	-0.214	0.152	**0.692**	0.14	-0.177	-0.116	-0.08	0.078
N sniff	0.44	**0.82**	0.042	0.081	0.372	0.208	**0.827**	0.022	-0.023	-0.016
N walk	**0.889**	0.225	0.124	-0.011	**0.822**	0.289	0.314	0.066	-0.034	-0.008
N total look	**0.793**	0.173	0.138	0.029	**0.728**	0.268	0.212	0.045	0.361	-0.017
Proportion sniff	0.335	**0.876**	0.048	0.085	0.264	0.16	**0.874**	-0.01	0.001	-0.017
Proportion walk	**0.914**	0.126	0.158	-0.063	**0.808**	0.187	0.327	0.127	-0.089	-0.026
Proportion look	**-0.849**	-0.43	-0.146	0.015	-0.49	**-0.625**	-0.391	-0.017	0.123	0.048
Latency to approach mirror (sec.)	-	-	-	-	-0.327	**-0.608**	-0.197	-0.066	-0.171	-0.082
Proportion spent at mirror	-	-	-	-	0.14	**0.841**	-0.06	-0.064	0.155	-0.018
N scratch mirror	-	-	-	-	0.426	0.763	0.206	-0.08	-0.097	-0.029
Proportion scratch mirror	-	-	-	-	0.169	**0.877**	0.168	-0.032	-0.11	-0.039
Percent variance explained	42.518	9.896	8.867	7.563	40.351	9.178	6.963	6.555	6.178	5.715
Total variance	42.518	52.413	61.28	68.843	40.351	49.528	56.491	63.047	69.224	74.939

Supplementary Table 2.

Traits, random effects, deviance information criteria (DIC), and delta (Δ) DIC for docility, activity and exploration in OF and MIS tests, and sociability. Random effects were removed from the full model (Individual + Maternal effect + Year) one at a time and delta DIC was calculated. Random effects with delta DIC greater than 2 were considered to make the model significantly better.

Trait	Random Effects	DIC	Δ DIC
Docility	Individual+Maternal+Year	17961.67	-
Docility	Maternal+Year	17965.6	3.93
Docility	Individual+Year	17972.98	11.31
Docility	Individual+Maternal	18020.3	58.63
Activity (OF1)	Individual+Maternal+Year	1125.758	-
Activity (OF1)	Maternal+Year	1137.434	11.676
Activity (OF1)	Individual+Year	1122.844	-2.914
Activity (OF1)	Individual+Maternal	1142.253	16.495
Exploration (OF2)	Individual+Maternal+Year	1310.282	-
Exploration (OF2)	Maternal+Year	1310.747	0.465
Exploration (OF2)	Individual+Year	1310.639	0.357
Exploration (OF2)	Individual+Maternal	1336.333	26.051
Activity (MIS1)	Individual+Maternal+Year	1195.528	-
Activity (MIS1)	Maternal+Year	1193.974	-1.554
Activity (MIS1)	Individual+Year	1194.957	-0.571
Activity (MIS1)	Individual+Maternal	1216.273	20.745
Sociability (MIS2)	Individual+Maternal+Year	1304.16	-
Sociability (MIS2)	Maternal+Year	1306.103	1.943
Sociability (MIS2)	Individual+Year	1308.175	4.015
Sociability (MIS2)	Individual+Maternal	1303.937	-0.223
Exploration (MIS3)	Individual+Maternal+Year	1352.081	-
Exploration (MIS3)	Maternal+Year	1356.61	4.529
Exploration (MIS3)	Individual+Year	1351.001	-1.08
Exploration (MIS3)	Individual+Maternal	1363.379	11.298

Supplementary Table 3.

Phenotypic variances, covariances, and correlations (P-matrix) of yellow-bellied marmot personality traits. Variances are present on the diagonal, the upper triangle contains the correlations, and the bottom triangle the covariances. Correlations and covariances were considered significant if they were different than 0 (based on the Bayesian equivalent of a 95% confidence interval). Significant values are in bold. Non-significant correlations over |0.400| have an asterisk.

Trait	Docility	Activity (OF1)	Exploration (OF2)	Activity (MIS1)	Sociability (MIS2)	Exploration (MIS3)
Docility	0.203 (0.168 - 0.249)	**-0.301 (-0.571 to - 0.074)**	0.104 (-0.238 - 0.350)	-0.225 (-0.441 - 0.063)	0.006 (-0.347 - 0.226)	0.290 (-0.006 - 0.527)
Activity (OF1)	**-0.073 (-0.151 to - 0.011)**	0.279 (0.167 - 0.436)	-0.275 (-0.528 - 0.079)	**0.571 (0.303 - 0.741)**	**0.450 (0.206 - 0.712)**	0.033 (-0.327 - 0.394)
Exploration (OF2)	0.020 (-0.057 - 0.101)	-0.056 (-0.197 - 0.022)	0.285 (0.162 - 0.533)	-0.156 (-0.483 - 0.187)	0.033 (-0.327 - 0.394)	0.335 (-0.049 - 0.623)
Activity (MIS1)	-0.073 (-0.171 to 0.024)	**0.194 (0.083 - 0.393)**	-0.033 (-0.221 - 0.097)	0.675 (0.402 - 0.907)	0.483 (0.185 - 0.687)	-0.357 (-0.563 - 0.039)
Sociability (MIS2)	-0.002 (-0.005 to 0.144)	**0.125 (0.044 - 0.253)**	0.034 (-0.096 - 0.121)	**0.147 (0.059 - 0.333)**	0.216 (0.137 - 0.406)	-0.172 (-0.395 - 0.314)
Exploration (MIS3)	0.066 (-0.005 to 0.144)	-0.010 (-0.096 - 0.121)	0.129 (-0.020 - 0.246)	-0.166 (-0.300 -0.023)	-0.045 (-0.126 - 0.089)	0.326 (0.177 - 0.555)

Supplementary Table 4.

Additive genetic variances, covariances, and correlations (G-matrix) of yellow-bellied marmot personality traits with the equivalent 95% confidence interval in parentheses. Variances are present on the diagonal, the upper triangle contains the correlations, and the bottom triangle the covariances. Correlations and covariances were considered significant if they were different than 0 (based on the Bayesian equivalent of a 95% confidence interval). Significant values are in bold. Non-significant correlations over |0.400| have an asterisk.

Trait	Docility	Activity (OF1)	Exploration (OF2)	Activity (MIS1)	Sociability (MIS2)	Exploration (MIS3)
Docility	0.064 (0.023 - 0.109)	-0.408 (-0.731 - 0.212)	-0.023 (-0.632 - 0.488)	-0.064 (-0.688 - 0.457)	-0.223 (-0.460 - 0.561)	0.191 (-0.383 - 0.653)
Activity (OF1)	-0.030 (-0.098 - 0.028)	0.145 (0.028 - 0.261)	-0.139 (-0.736 - 0.377)	0.660 (-0.134 - 0.890)	**0.673** **(0.005 - 0.833)**	0.302 (-0.518 - 0.605)
Exploration (OF2)	-0.010 (-0.064 - 0.048)	-0.020 (-0.111 - 0.064)	0.089 (0.018 - 0.183)	-0.314 (-0.615 - 0.646)	-0.008 (-0.528 - 0.586)	0.271 (-0.377 - 0.720)
Activity (MIS1)	-0.022 (-0.109 - 0.068)	0.079 (-0.037 - 0.228)	-3.6e10-4 (-0.104 - 0.106)	0.173 (0.022 - 0.427)	0.549 (-0.233 - 0.833)	-0.277 (-0.838 - 0.342)
Sociability (MIS2)	0.007 (-0.044 - 0.068)	0.082 (-0.007 - 0.170)	-0.003 (-0.084 - 0.066)	0.063 (-0.038 - 0.217)	0.089 (0.021 - 0.177)	-0.085 (-0.657 - 0.480)
Exploration (MIS3)	0.013 (-0.040 - 0.074)	0.008 (-0.084 - 0.107)	0.025 (-0.054 - 0.125	-0.047 (-0.207 - 0.056)	-0.016 (-0.093 - 0.064)	0.121 (0.020-0.247)

Supplementary Table 5.

Permanent environment variances, covariances, and correlations of yellow-bellied marmot personality traits. Variances are present on the diagonal, the upper triangle contains the correlations, and the bottom triangle the covariances. Correlations and covariances were considered significant if they were different than 0 (based on the Bayesian equivalent of a 95% confidence interval). Significant values are in bold. Non-significant correlations over |0.400| have an asterisk.

Trait	Docility	Activity (OF1)	Exploration (OF2)	Activity (MIS1)	Sociability (MIS2)	Exploration (MIS3)
Docility	0.097 (0.064 - 0.134)	-0.497 (-0.724 - 0.103)	0.209 (-0.276 - 0.606)	-0.104 (-0.571 - 0.260)	-0.158 (-0.617 - 0.380)	**0.521 (0.070 - 0.806)**
Activity (OF1)	-0.035 (-0.080 - 0.016)	0.085 (0.023 - 0.166)	-0.493 (-0.752 - 0.104)	**0.641 (0.095 - 0.862)**	0.497 (-0.199 - 0.740)	-0.309 (-0.610 - 0.486)
Exploration (OF2)	0.028 (-0.030 - 0.089)	-0.056 (-0.138 - 0.021)	0.141 (0.023 - 0.283)	-0.322 (-0.747 - 0.272)	-0.132 (-0.557 - 0.568)	0.439 (-0.144 - 0.805)
Activity (MIS1)	-0.030 (-0.118 - 0.056)	**0.124 (0.006 - 0.269)**	-0.062 (-0.188 - 0.085)	0.411 (0.118 - 0.656)	0.539 (-0.057 - 0.820)	-0.624 (-0.799 - 0.196)
Sociability (MIS2)	-0.009 (-0.052 - 0.044)	0.026 (-0.015 - 0.082)	0.006 (-0.080 - 0.070)	0.107 (-0.018 - 0.223)	0.071 (0.020 - 0.144)	-0.120 (-0.492 - 0.570)
Exploration (MIS3)	-0.018 (-0.056 - 0.016)	0.006 (-0.047 - 0.060)	0.064 (-0.047 - 0.165)	-0.002 (-0.058 - 0.056)	0.004 (-0.068 - 0.086)	0.140 (0.024 - 0.278)

Supplementary Table 6.

Maternal effect variances, covariances, and correlations of yellow-bellied marmot personality traits. Variances are present on the diagonal, the upper triangle contains the correlations, and the bottom triangle the covariances. Correlations and covariances were considered significant if the were different than 0 (based on the Bayesian equivalent of a 95% confidence interval). Significant values are in bold. Non-significant correlations over |0.400| have an asterisk.

Trait	Docility	Activity (OF1)	Exploration (OF2)	Activity (MIS1)	Sociability (MIS2)	Exploration (MIS3)
Docility	0.046 (0.022 - 0.071)	-0.351 (-0.673 - 0.232)	-0.021 (-0.485 - 0.478)	-0.329 (-0.688 - 0.207)	-0.105 (-0.589 - 0.306)	-0.247 (-0.620 - 0.271)
Activity (OF1)	-0.017 (-0.050 - 0.015)	0.058 (0.018 - 0.114)	-0.028 (-0.585 - 0.486)	0.314 (-0.184 - 0.764)	0.126 (-0.275 - 0.754)	0.039 (-0.427 - 0.590)
Exploration (OF2)	-0.001 (-0.038 - 0.042)	-0.003 (-0.052 - 0.049)	0.089 (0.018 - 0.183)	0.044 (-0.573 - 0.555)	0.342 (-0.529 - 0.647)	0.230 (-0.380 - 0.708)
Activity (MIS1)	-0.022 (-0.062 - 0.018)	0.027 (-0.016 - 0.089)	-3.6e10-4 (-0.104 - 0.106)	0.070 (0.014 - 0.153)	0.200 (-0.406 - 0.741)	0.120 (-0.563 - 0.548)
Sociability (MIS2)	-0.015 (-0.052 - 0.023)	0.027 (-0.031 - 0.100)	0.012 (-0.068 - 0.088)	0.019 (-0.040 - 0.104)	0.105 (0.019 - 0.214)	-0.333 (-0.642 0.509)
Exploration (MIS3)	-0.018 (-0.056 - 0.016)	0.006 (-0.047 - 0.060)	0.018 (-0.047 - 0.088)	-0.002 (-0.058 - 0.056)	0.079 (-1.587 - 1.429)	1.460 (0.024 - 3.206)

References

Armitage, K.B. 2010. Individual fitness, social behavior, and population dynamics of yellow-bellied marmots. *Ecol. Place Contrib. Place-Based Res. Ecol. Underst.* 132–154.

Armitage, K.B. 1986. Individuality, social behavior, and reproductive success in yellow-bellied marmots. *Ecology* **67**: 1186–1193.

Bell, A.M. 2005. Behavioural differences between individuals and two populations of stickleback (*Gasterosteus aculeatus*). *J. Evol. Biol.* **18**: 464–473.

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

Biro, P.A. & Stamps, J.A. 2008. Are animal personality traits linked to life-history productivity? *Trends Ecol. Evol.* **23**: 361–368.

Blumstein, D.T. & Daniel, J.C. 2007. *Quantifying behavior the JWatcher way*. Sinauer Associates Inc.

Blumstein, D.T., Ozgul, A., Yovovich, V., Van Vuren, D.H. & Armitage, K.B. 2006. Effect of predation risk on the presence and persistence of yellow-bellied marmot (*Marmota flaviventris*) colonies. *J. Zool.* **270**: 132–138.

Blumstein, D.T., Wey, T.W. & Tang, K. 2009. A test of the social cohesion hypothesis: interactive female marmots remain at home. *Proc. R. Soc. B Biol. Sci.* **276**: 3007–3012.

Boon, A.K., Réale, D. & Boutin, S. 2007. The interaction between personality, offspring fitness and food abundance in North American red squirrels. *Ecol. Lett.* **10**: 1094–1104.

Carter, A.J., Feeney, W.E., Marshall, H.H., Cowlishaw, G. & Heinsohn, R. 2013. Animal personality: what are behavioural ecologists measuring? *Biol. Rev.* **88**: 465–475.

Dingemanse, N.J., Wright, J., Kazem, A.J., Thomas, D.K., Hickling, R. & Dawnay, N. 2007. Behavioural syndromes differ predictably between 12 populations of three-spined stickleback. *J. Anim. Ecol.* **76**: 1128–1138.

Dochtermann, N.A. 2010. Behavioral syndromes: carryover effects, false discovery rates, and a priori hypotheses. *Behav. Ecol.* **21**: 437–439.

Falconer, D.S. & Mackay T.F.C.1996. *Introduction to quantitative genetics.* (4th ed) Longmans Green, Harlow, Essex, UK.

Frase, B.A. & Hoffmann, R.S. 1980. Marmota flaviventris. *Mamm. Species* 1–8.

Gosling, S.D. 2001. From mice to men: what can we learn about personality from animal research? *Psychol. Bull.* **127**: 45.

Hadfield, J.D. 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *J. Stat. Softw.* **33**: 1–22.

Kalinowski, S.T., Taper, M.L. & Marshall, T.C. 2007. Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* **16**: 1099–1106.

Krause, J., James, R. & Croft, D.P. 2010. Personality in the context of social networks. *Philos. Trans. R. Soc. B Biol. Sci.* **365**: 4099–4106.

Kruuk, L.E.B. 2004. Estimating genetic parameters in natural populations using the "animal model." *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **359**: 873–890.

Kruuk, L.E.B., Clutton-Brock, T.H., Slate, J., Pemberton, J.M., Brotherstone, S. & Guinness, F.E. 2000. Heritability of fitness in a wild mammal population. *Proc. Natl. Acad. Sci.* **97**: 698–703.

Lande, R. & Arnold, S.J. 1983. The measurement of selection on correlated characters. *Evolution* **37**: 1210.

Lea, A.J., Blumstein, D.T., Wey, T.W. & Martin, J.G. 2010. Heritable victimization and the benefits of agonistic relationships. *Proc. Natl. Acad. Sci.* **107**: 21587–21592.

Lynch, M. & Walsh, B. 1998. *Genetics and analysis of quantitative traits*. Sinauer Associates Inc.

Olson, L.E. & Blumstein, D.T. 2010. Applying the coalitionary-traits metric: sociality without cooperation in male yellow-bellied marmots. *Behav. Ecol.* **21**: 957–965.

Petelle, M. & Blumstein, D. 2014. A critical evaluation of subjective ratings: Unacquainted observers can reliably assess certain personality trait. *Curr. Zool.* **60**: 162–169.

Petelle, M.B., McCoy, D.E., Alejandro, V., Martin, J.G. & Blumstein, D.T. 2013. Development of boldness and docility in yellow-bellied marmots. *Anim. Behav.* **86**: 1147–1154.

Räsänen, K. & Kruuk, L.E.B. 2007. Maternal effects and evolution at ecological time-scales. *Funct. Ecol.* **21**: 408–421.

RDevelopment, C. 2012. TEAM 2009: R: A Language and Environment for Statistical Computing. *Vienna Austria Internet http://www.R-Proj.org*.

Réale, D., Gallant, B.Y., Leblanc, M. & Festa-Bianchet, M. 2000. Consistency of temperament in bighorn ewes and correlates with behaviour and life history. *Anim. Behav.* **60**: 589–597.

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

Réale, D., Martin, J., Coltman, D.W., Poissant, J. & Festa-Bianchet, M. 2009. Male personality, life-history strategies and reproductive success in a promiscuous mammal. *J. Evol. Biol.* **22**: 1599–1607.

Réale, D., Reader, S.M., Sol, D., McDougall, P.T. & Dingemanse, N.J. 2007. Integrating animal temperament within ecology and evolution. *Biol. Rev.* **82**: 291–318.

Reinhold, K. 2002. Maternal effects and the evolution of behavioral and morphological characters: A literature review indicates the importance of extended maternal care. *J. Hered.* **93**: 400–405.

Roff, D.A. & Fairbairn, D.J. 2007. The evolution of trade-offs: where are we? *J. Evol. Biol.* **20**: 433–447.

Sih, A., Bell, A.M., Johnson, J.C. & Ziemba, R.E. 2004. Behavioral syndromes: an integrative overview. *Q. Rev. Biol.* **79**: 241–277.

Sinn, D.L., Apiolaza, L.A. & Moltschaniwskyj, N.A. 2006. Heritability and fitness-related consequences of squid personality traits. *J. Evol. Biol.* **19**: 1437–1447.

Smith, B.R. & Blumstein, D.T. 2008. Fitness consequences of personality: a meta-analysis. *Behav. Ecol.* **19**: 448–455.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & Van Der Linde, A. 2002. Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64**: 583–639.

Stirling, D.G., Réale, D. & Roff, D.A. 2002. Selection, structure and the heritability of behaviour. *J. Evol. Biol.* **15**: 277–289.

Taylor, R.W., Boon, A.K., Dantzer, B., Reale, D., Humphries, M.M., Boutin, S., *et al.* 2012. Low heritabilities, but genetic and maternal correlations between red squirrel behaviours. *J. Evol. Biol.* **25**: 614–624.

Van Oers, K., de Jong, G., van Noordwijk, A.J., Kempenaers, B. & Drent, P.J. 2005. Contribution of genetics to the study of animal personalities: a review of case studies. *Behaviour* **142**: 9–10.

Weaver, I.C.G., Cervoni, N., Champagne, F.A., D'Alessio, A.C., Sharma, S., Seckl, J.R., *et al.* 2004. Epigenetic programming by maternal behavior. *Nat. Neurosci.* **7**: 847–854.

Wey, T., Blumstein, D.T., Shen, W., & Jordán F. 2008. Social network analysis of animal behaviour: a promising tool for the stuy of sociality. *Anim. Behav.* **75**: 333–344.

Wilson, A.J., Reale, D., Clements, M.N., Morrissey, M.M., Postma, E., Walling, C.A., *et al.* 2010. An ecologist's guide to the animal model. *J. Anim. Ecol.* **79**: 13–26.

Wilson, D.S. 1998. Adaptive individual differences within single populations. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **353**: 199–205.

Wolf, M., van Doorn, G.S., Leimar, O. & Weissing, F.J. 2007. Life-history trade-offs favour the evolution of animal personalities. *Nature* **447**: 581–584.