MULTI-LEVEL VARIATION IN LABILE CHARACTERS: ADAPTIVE CAUSES AND EVOLUTIONARY CONSEQUENCES

Dissertation

der Fakultät für Biologie

der Ludwigs-Maximilians-Universität München

zur Erlangung des Doktorgrades der

Naturwissenschaften Dr. rer. nat.

durchgeführt am

Max-Planck-Institut für Ornithologie

Seewiesen

vorgelegt von

Yimen Gerardo Araya Ajoy

München 2015

Erstgutachter: Prof. Dr. **Niels Dingemanse**

Zweitgutachter: Prof. Dr. **Bart Kempenaers**

Tag der Abgabe: **15.04.2015**

Tag der mündlichen Prüfung: **06.07.2015**

Table of Contents

Summary

Labile characters, like behaviors, are phenotypes that are expressed repeatedly in the life of an individual. These types of characters allow individuals to adjust their phenotype to various levels of environmental variation, and therefore play a key role in the evolutionary process. Labile phenotypes are distinct because of their multi-level nature; individuals can differ in their average phenotypic expression (causing among-individual variation), but they can also vary their phenotype in each expression (causing within-individual variation). In order to understand the role of labile characters in the evolutionary process it is necessary to acknowledge that variation at each level is caused by different processes. Variation at the among-individual level is caused by genetic or environmental differences having a permanent effect on an individual's phenotype, whereas variation at the within-individual level is caused by an individual's adjustment of its phenotype to a changing environment. The implications of these multi layered effects in the expression of labile characters have been acknowledged by different fields of evolutionary ecology, but major areas of evolutionary research do not fully incorporated this idea. The general aim of my thesis was to fully integrate this multi-level nature in the study of the adaptive causes and evolutionary consequences of variation in labile characters. My thesis is composed of five chapters: the first three are conceptual and methodological works aimed at integrating the multi-level nature of labile characters into already existing evolutionary frameworks. The last two chapters describe, as a worked example, how the different levels of variation and covariation between (labile) fertilization related traits affect the evolution of the alternative reproductive strategies in a wild passerine bird (the great tit).

The first chapter is a conceptual work focusing on how to define and statistically characterize behavioral characters. We argue that behavioral characters can be studied using the "evolutionary character concept". This framework was developed to study characters that only vary among individuals (i.e. "fixed characters"); therefore we extended this framework to include characters that also vary within-individuals. The second chapter of the thesis is a methodological work where we proposed a way to quantify multi-level variation in reaction norms, which allows the estimation of repeatability of plasticity. Behavioral ecologists have recently developed theory predicting the ecological conditions where repeatable vs.

non-repeatable variation in phenotypic plasticity should evolve. However, there was no methodological framework to estimate repeatability of plasticity. Therefore, we proposed a study design and mixed effect model structure to estimate repeatability of plasticity. To help researchers use the proposed methodology, we developed an R simulation package to estimate bias, precision and accuracy for different sampling designs. The third chapter is an opinion paper that urges researchers to combine theory and methods developed in behavioral ecology and quantitative genetics to study phenotypic variation in a social context. Quantitative geneticists have developed a framework to study social evolution aimed at predicting the evolutionary response to selection of traits affected by the phenotypes of other individuals (the “social environment”). Phenotypes expressed in a social context, also called interactive phenotypes, exhibit a particular evolutionary dynamic because their environmental component is composed of genes and can thus evolve. Despite that fact that the effects of the social environment are commonly mediated by labile characters, this social evolution framework has not fully considered the multi-level nature of labile characters. Therefore, for chapter three we integrated the multi-level nature of labile characters in this social evolution framework.

The final two chapters focus, as a worked example, on within-pair and extra-pair reproductive behavior in great tits. For these chapters, we utilized the theoretical and methodological developments of the previous chapters to study the sources of evolutionary constraints on alternative fertilization routes in male great tits. One of the chapters has a more evolutionary perspective, while the other applies a more behavioral ecology view point. In chapter four we studied male extra-pair and within-pair reproduction as interactive phenotypes that are affected by the phenotypes of both the male and the female member of great tit breeding pairs. We showed that male fertilization strategies depend heavily on the phenotype of their female. This social environment effect should influence the evolutionary response to selection of male fertilization strategies, and could partly explain evolutionary stasis, observed in natural populations, in traits so closely linked to fitness. In chapter four we also studied whether trade-offs among- or within-individuals can constrain the phenotypic evolution of male alternative reproductive strategies. We showed that among-male trade-offs between within-pair and extra-pair reproduction could also be a source of evolutionary constrain. In chapter five, we corroborated the existence of trade-offs between alternative

reproductive routes by studying whether within-pair and extra-pair fertilizations are obtained at the same time, allowing for the possibility of a trade-off between the two. We found that a male's extra-pair fertilization success is actually higher when it constrains his ability to secure within-pair fertilizations. This result is consistent with our finding that there is indeed a trade-off between extra-pair and within-pair reproduction in this species. The empirical works in this thesis highlight the importance of the social environment as a source of phenotypic variation in the expression of labile traits. But more generally, from the works in this thesis, we can conclude that to fully understand the role of labile characters in the evolutionary process it is necessary to acknowledge their multi-level nature.

General Introduction

What role do behaviors play in the evolutionary process? Evolutionary biologists have argued that they are key players on the evolutionary stage (Hamilton 1964; Maynard-Smith & Price 1973; West-Eberhard 1979). Behavior has been hypothesized to be a pacemaker in the rate of evolution, involved in the acceleration of the evolutionary process but also a cause of phenotypic stasis (Huey et al. 2003; Duckworth 2008; Wilson 2013). Behavior could mediate both evolutionary processes, but to understand why and when it will increase or decrease the rate of phenotypic evolution, it is necessary to acknowledge its multi-level nature. Phenotypic variation is organized in a hierarchical way; variation can exist among-species, within-species among-populations, and within-populations among-individuals (Figure 1). In the case of characters that are repeatedly expressed throughout the life of an individual (labile characters), variation can also exist within-individuals among-expressions (Westneat et al. 2014). Labile characters, like behavior, are very important from an evolutionary perspective because they provide individuals with the means to adapt to a constantly varying environment. Behavior's evolutionary importance has been mostly attributed to this labile nature (capacity to vary within-individuals; Westneat & Fox 2010), but the adaptive nature of differences between individuals in their behavior (among-individual variation) has also been studied from an adaptive perspective (Dingemanse & Wolf 2010). Recently, there have been repeated calls for integrating both the among- and within-individual levels of variation in behavior in an evolutionary context (Nussey et al. 2007; Dingemanse et al. 2010; Westneat et al. 2014). The need to integrate different levels of phenotypic variation to understand evolutionary processes has been clear since Darwin and Wallace connected among-species and among-individual variation through the process of natural selection (Darwin 1859). Therefore, in this thesis we aimed to increase the integration of “lower” levels of variation in the study of phenotypic variation from an evolutionary perspective. We studied the adaptive causes of variation among- and within-individuals, how these levels of variation relate to the way individuals respond to environmental variation, the connection between these levels of variation, and how populations may respond to selection.

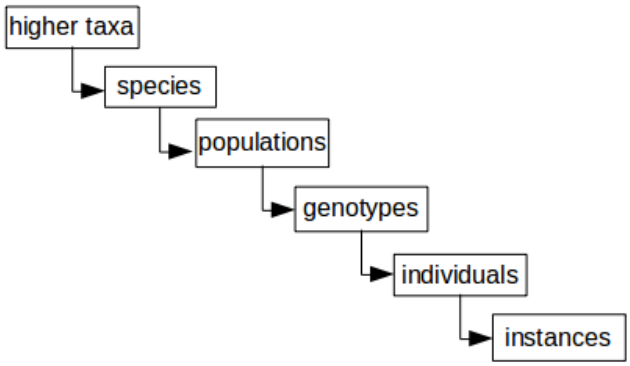

***Figure 1**. Schematic representation of the hierarchical organization of phenotypic variance (modified from Weastneat et. al., 2014). Directional arrows indicate that replicates of the next level are nested within the upper level.*

Multi-level variation of behavior

The adaptive nature of within-individual variation in behavior (individual plasticity) has long been of interest to behavioral ecologists (Westneat & Fox 2010). Traditionally, behavioral ecology research has focused on how individuals adjust their behavior to match environmental conditions. Based on optimality theory, behavioral ecologists placed phenotypic plasticity (within-individual variation) in a central position in their adaptive explanations of behavioral variation (Krebs & Davies 1997). For instance, birds are known to adjust their level of aggressiveness depending on the costs and benefits of an aggressive interaction (Enquist & Leimar 1983) and females optimize their clutch size in response to yearly variation in density (Both 1998). More recently, among-individual variation in behavior has also become the focus of behavioral ecologists (Dall et al. 2004; Réale et al. 2007). Within the field of 'animal personality', consistent individual differences across time and contexts have been documented in a wide range of behaviors and taxa (Bell et al. 2009), while theoretical and empirical studies have expanded our understanding of its adaptive nature (e.g Wolf et al. 2007; McNamara et al. 2009; reviewed by Dingemanse & Wolf 2010). For example, behavioral differences between individuals can be adaptive because of the benefits of reduced competition during social interactions (Bergmüller & Taborsky 2010; Montiglio et al. 2013) and increased efficiency due to task specialization (Laskowski & Pruitt 2014). Behavioral ecologists are now studying among- and within-individual variation in

behavior in conjunction within a single (reaction norm) framework (Dingemanse et al. 2010). Reaction norms are functions that describe the dependency of a phenotype on the environment (Schlichting & Pigliucci 1998). Behavioral reaction norms are characterized by an intercept, representing an individual's average level of behavior, and a slope that represents its degree of phenotypic plasticity. The plastic response of individuals to the environment (slope) causes within-individual variation (Figure 2) while variation in the elevation of the reaction norms (intercepts) will reflect consistent among-individual differences across the environmental gradient (Figure 2a). This reaction norm framework has also made it possible to study the interactions between these levels (Nussey et al. 2007; Dingemanse et al. 2010). It is now widely documented that there are differences between individuals in how they plastically respond to the environment (Figure 2b; van de Pol 2012). In other words, this phenomenon can be described as among-individual variation in a source of within-individual variation. The adaptive nature of this type of variation has also been studied theoretically (Dingemanse & Wolf 2013) and empirically (Mathot et al. 2011). Currently, researchers studying the evolutionary ecology of behavioral variation are developing theory and methodologies to study different ways in which levels of variation can interact (Biro & Adriaenssens 2013; Cleasby et al. 2014; Westneat et al. 2014).

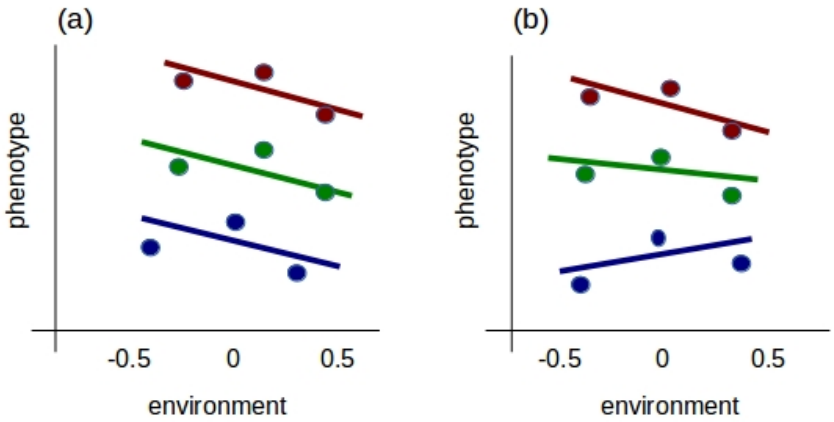

Figure 2. *Linear reaction norms (modified from Dingemanse et. al., 2010), where each color represents a different individual. (a) Depicts a situation where individuals respond in the same way to a particular environmental gradient. (b) Represents a situation where individuals respond differently to the same environmental gradient. The plastic response to the environment generates within-individual variation in phenotypic expression. Differences in the elevation of each of the lines represents variation among-individuals.*

Multi-level variation and population response to natural selection

The evolutionary response to natural selection of a population will depend on the relative contribution of the among- and within-individual variation to the total phenotypic variance in a population (Lynch & Walsh 1998). Among-individual variance due to genetic differences is the raw material for natural selection to act upon and adaptive phenotypic evolution to proceed, whereas within-individual variation in the form of phenotypic plasticity allows individuals to adjust their phenotype to the environment without genetic change. How the different levels of phenotypic variation are related to the evolutionary response to selection by a population is also determined by the sources and levels of covariation between traits and fitness. Covariance between labile characters and fitness can be caused by different processes, result in covariances at different levels, and have different evolutionary repercussions. Covariation between behavior and fitness can be the result of a correlated plastic response, of both behavior and fitness, to the same environmental gradient (i.e. environmental pleiotropy; Figure 3a). For example, an increase in food availability could increase individual's aggressiveness and number of offspring. This will cause an environmental correlation between behavior and fitness (Figure 3a). This can also be referred to as a within-individual correlation when studying labile traits, because the pattern is caused by the changes within-individuals due to environmental effects—it is not caused by consistent individual differences. This environmental correlation between trait and fitness will not result in adaptive phenotypic evolution (Sheldon et al. 2003), and despite covariation between behavior and fitness, phenotypic stasis will be observed (Merilä et al. 2001). Responses to selection are only expected if the covariance between behavior and fitness is at the among-individual level (Figure 3b) and underpinned by an additive genetic covariance (Lynch & Walsh 1998). In this scenario, individuals that are always more aggressive always have more offspring. Therefore to determine the role of behaviors in the evolutionary processes, it is necessary to acknowledge that covariation between behavior and fitness can occur at different levels and be caused by different processes and therefore have different evolutionary consequences.

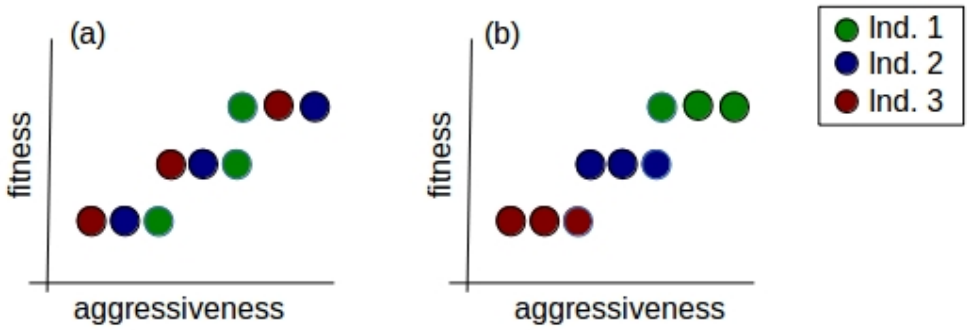

Figure 3. *Different ways in which fitness and labile characters can be correlated; colors represent different individuals. (a) Correlated plastic response to an environmental gradient results in a correlation within individuals between trait and the fitness. In instances that individuals were more aggressive they also sired more offspring, but individuals were not consistent in their aggressiveness across instances. (b) The correlation between fitness and the labile traits is at the among-individual level. Individuals that were consistently more aggressive consistently sired more offspring.*

Multi-level variation and the social environment

Most behaviors are expressed in a social context, and social interactions are key determinants of population level processes. Despite the role that behavioral characters play in mediating social interactions, the multi-level nature of social behaviors has not been fully incorporated in a social evolution framework. Theory and methods developed in quantitative genetics may prove useful for bridging the gap between the multi-level nature of behavior and social evolution (McGlothlin & Brodie 2009). This is because quantitative geneticists are interested in predicting evolutionary responses to selection, and this explicitly requires the partitioning of phenotypic variation into genetic (among-individual) versus environmental (within-individual) components (Falconer & Mackay 1996). Specifically, indirect genetic effects and social selection theory center upon the proposition that during social interactions the environmental component of one individual is the phenotype of another (Moore et al. 1997). This can have major evolutionary repercussions because the (social) environment can be heritable and potentially also evolve (McGlothlin et al. 2010). This type of evolutionary dynamic can increase or decrease the rate of phenotypic evolution depending on the nature of the social interactions (Wolf 2003; Wilson 2013). For example, in mew gulls a conflict of interest between the sexes over the timing of reproduction constrains the optimal reproduction time for each sex (Brommer & Rattiste 2008). In mice, competition among males can explain phenotypic stasis in traits like dominance, due to indirect genetic effects (Wilson et al. 2011). Furthermore, the evolution of exaggerated sexual ornaments can be explained by these

social evolutionary dynamics (Westneat 2012). In all of these scenarios, the effect of the social environment on the phenotype of one individual is mediated by phenotypic plasticity. During social interactions, phenotypic plasticity as a function of phenotypes expressed by conspecifics (i.e. social responsiveness) represents the link between the genotypes in the social environment and the phenotypes expressed by an individual. Combining the social evolutionary framework with behavioral ecology theory concerning the multi-level nature of behavior and plasticity, will deepen our understanding of the role of labile characters in social evolution.

The role of social behaviors in the rate of phenotypic evolution will ultimately depend upon their relationship with life-history traits and consequently fitness (Morrissey 2014). Life-history traits are those traits that affect the survival and reproductive potential of individuals (Roff 1992). In species where individuals reproduce several times in their life, some life-history traits vary among- and within-individuals (Browne & McCleery 2007). Moreover, life- history traits are commonly affected by the phenotypes of other individuals (and thus other genotypes), which generates a similar type of evolutionary dynamic as in social behaviors (Wilson 2013). Life-history traits are closely linked to fitness, and therefore covariation at different levels between social behaviors and life-history traits will determine the evolutionary trajectories of populations (Stearns 1989). To understand why and when social behaviors will increase or decrease the rate of phenotypic evolution, it is necessary to study social behaviors and their relationship with life-history traits in a unified framework that acknowledges the different levels and sources of variation on both life-history traits and social behaviors.

Variation in social behaviors, life history traits, and extra-pair reproduction

One such relationship between life-history traits and social behaviors are male alternative reproductive strategies (Gross 1996), where males use a variety of behaviors to mediate their investment in alternative reproductive routes. In socially monogamous birds, extra-pair reproduction results in a polygynandrous reproductive system that creates alternative routes to fertilization success for both males and females (Kvarnemo & Simmons 2013). Socially monogamous males can achieve successful fertilizations via three routes: mating with highly fecund females, avoiding within-pair paternity loss, and/or seeking

extra-pair copulations (Webster et al. 1995). Variation in these alternative reproductive routes will be partly due to among-male variation in behavior, among-female variation in behavior, and environmentally induced variation (Westneat & Stewart 2003). Therefore, fertilization routes not only vary at multiple-levels, but are also affected by the behavioral phenotype of various individuals. Studying variation among behavioral traits and alternative reproductive routes within a unified framework should give a better understanding of how their covariation shapes and is shaped by the mating system characteristics of socially monogamous species.

Thesis outline

The general goal of my thesis was to study how (co) variation of labile characters among- and within-individuals relates to the way by which individuals respond to environmental variation, and how this may influence population response to natural selection. As a specific case of this phenomenon, we studied how male great tits mediate investment in different reproductive routes using aggressive behaviors, and how the different sources of variation and covariation between within- and extra-pair reproduction constrain the evolution of alternative reproductive strategies.

Before trying to understand the role of a behavioral character (e.g., aggressiveness) in any biological process (e.g., reproduction), it is necessary to define and properly quantify the behavioral character in question. What is a character and how can we measure it? These questions are of key importance when studying any phenotypic character from an evolutionary perspective. In **chapter one,** we set out to answer these questions for labile characters. Biologists often study phenotypic evolution while assuming that phenotypes consist of a set of quasi-independent units or parts that have been shaped by selection to accomplish a particular function (Wagner 2001). Consequently, the success of any evolutionary research agenda depends to a large degree on whether such functional units have been properly characterized. Despite the importance of labile characters as mediators between organisms and their environment (Krebs & Davies 1997; Westneat & Fox 2010), theory about what defines an evolutionary character has been developed to study characters that only vary among-individuals (“fixed characters”). Thus, the conceptual approaches to understand what a character constitutes have neglected the plastic nature of behavior. Therefore, we propose a concept of “behavioral characters” that integrates both variation

among-individuals (“personality”) and within-individuals (“individual plasticity”), and further propose a corresponding statistical methodology to test whether observed behaviors should be considered expressions of a hypothesized evolutionary character.

The conceptual and methodological framework we proposed is generally applicable to any labile character, therefore we chose to present the study of aggressiveness in great tits as an empirical test of a general framework. Our approach hinged on the notion that a behavioral character should be viewed as a latent variable underlying the expression of functionally related traits. For example, during aggressive interactions, male great tits express a suite of behaviors (sing, alarm, attack) that jointly execute a specific function: displacing intruders. We defined behavioral characters (aggressiveness) as the common neurobiological and physiological mechanisms that allow the functional coherence of the (aggressive) display. If the behavioral character of interest is actually underpinning the expression of the measured behaviors, it should cause a pattern of among- and within-individual covariance between behaviors, that will allow them to be used as functional unit within and across contexts. To empirically test whether the behaviors that we observed in the aggressive displays of great tits were indeed expressions of the behavioral character (aggressiveness), we confronted males in our population with a set of simulated territorial intrusions (Figure 3). Each male was subjected to a standardized simulated territorial intrusion in two different contexts: a situation with a high risk of a conspecific intrusion and another with a low risk of intrusion. Using multivariate mixed-effect models and structural equation modeling, we showed that the patterns of covariation among- and within-individuals of the different behaviors supported the hypothesis that the behaviors were used as a functional unit and were expressions of the behavioral character “aggressiveness”. This study informed us about the best measure of aggressiveness to incorporate in further analysis aimed at studying the relation between among- and within- individual variation in aggressiveness and the alternative routes to fertilization success of male great tits.

In chapter one, we studied the multi-level nature of phenotypic expression and how to incorporate it in the study of evolutionary characters. Another way by which the multi-level nature of labile characters is manifested is by the plastic response of individuals to the environment.

Figure 3. *Photograph of the experimental setup showing a taxidermic mount (protected by a green wire mesh) one meter away from the focal bird's nest box. The focal bird (on top of the wire mesh) is highly aggressive. Photograph by Jan Wijmenga*

Recently, among-individual variation in phenotypic plasticity has been studied by evolutionary ecologists. Plastic responses can also vary within-individuals, but this level of variation has not been studied. Therefore, in **chapter two** we detailed a study design and statistical approach—based on repeated measures and multi-level random regression modeling—that enables the study of variation in phenotypic plasticity at different hierarchical levels (such as among- and within-individuals). This methodology applies to labile characters that respond plastically to environmental gradients that individuals encounter several times throughout their life. Variation within-individuals in their plastic response to an environmental gradient may be caused by the dependency of their plastic response on a second environmental gradient. For example, red knots vary their vigilance behavior depending on the level of predation risk every day, but the plastic response towards predation risk could depend on daily variation in the size of their flock. The ability of individuals to regulate their plastic response to match a multivariate environment is very relevant from an adaptive perspective and is probably common in nature. Moreover, recent theoretical models have specific predictions regarding the ecological conditions where repeatable vs. non-repeatable variation in phenotypic plasticity should emerge. The quantification of the different levels of

variation in phenotypic plasticity allows the estimation of its repeatability, which is important to empirically determine when and why there is repeatable variation in plasticity.

The aim of **chapter three** was to develop a framework that might be applied to usefully integrate social evolution and the multi-level nature of behavior and any labile character. In this chapter we urged researchers to combine theory and methods developed in behavioral ecology and quantitative genetics to study, within a unified framework, the multi-level nature of labile phenotypes (e.g., life-history traits and behavior) and how social interactions may shape these levels of variation. In this chapter, we first reviewed the different experimental designs and statistical methodologies that will enable researchers to study the effects of the social environment on the different levels of phenotypic variation in labile traits. We then detailed, which biological hypotheses can be answered with these methods and further propose future areas of research that can be addressed with the proposed approach.

In **chapter four** we used the approaches proposed in chapter three combined with life-history theory to study, as a worked examle, the role of aggressiveness and the social environment in the fertilization strategies of male great tits. Extra-pair reproduction in socially monogamous species creates alternative routes to male fertilization success. In natural populations, variation in these routes is pervasive, which is puzzling, because selection should deplete variation in traits closely linked to fitness. In this chapter, we determined whether social environment effects (due to phenotypes of social mates) and the existence of trade-offs between fertilization routes, can explain the maintenance of variation and phenotypic stasis of traits linked to fertilization success and therefore fitness. Empirically addressing these type of questions is challenging in wild populations because trade-offs are often hidden at specific hierarchical levels of phenotypic organization (i.e., within-individuals as opposed to among-individuals: Stearns 1989), while male fertilization routes are simultaneously affected by the phenotype of both members of the social pair. We addressed this issue by using a (co)variance partitioning approach to study alternative routes to fertilization success in male great tits (*Parus major*). We first studied male siring routes as “interactive phenotypes” arising from phenotypic contributions of both members of the social pair. We then studied the relationships between the different fertilization routes to determine whether covariation within- or among-individual males supported the existence of a trade-off. As a general conclusion for this chapter, we found that both male and female individual-level

phenotypic attributes contribute to male fertilization success and trade-offs between the different routes may help maintain among-individual variance in alternative pathways to male fertilization success. In **chapter five,** we studied more in depth the trade-off between extra-pair paternity gain and within-pair paternity loss. We addressed whether the timing of extra-pair fertilizations may interfere with a male's ability to secure within-pair fertilizations. We found that spill-over effects of male within-pair fertilization behavior affects extra-pair fertilizations, causing both within-pair and extra-pair fertilizations to be achieved at the same time. This supports our results about the existence of a trade-off between extra-pair paternity gain and within-pair paternity loss.

Study system

For the empirical components of the thesis, we studied great tit (*Parus major*) populations breeding in nest boxes. The great tit is a non-migratory passerine bird from the family Paridae. It is the most widespread species of its genus and is common throughout Europe in any sort of woodland (Svensson 1992). The species readily breed in nest boxes between March and June when male great tits defend their territories (Krebs 1982). This bird is socially monogamous and provides bi-parental care to its young (Kölliker et al. 2000), but commonly engages in extra-pair reproduction (Brommer et al. 2007). We studied variation in aggressiveness and male alternative fertilization routes in 12 nest box plots of great tits (Figure 4), established in 2009 in Southern Germany in an area of approximately 120 ha (Bavarian Landkreis Starnberg; 47º 58' N, 11º 14' E). Each plot consisted of a regular grid of 50 boxes, with 50 meters between adjacent boxes. From April onwards, boxes were checked twice a week to determine lay date (back-calculated assuming that one egg was laid per day), onset of incubation, and clutch size. When the nestlings were 6 days old, a blood sample was taken and they were marked with an aluminum ring; any unhatched eggs or deceased nestlings were collected. Parents were caught with a spring trap in the nest box on the next day, measured, bled, and marked with a unique combination of rings if caught for the first time.

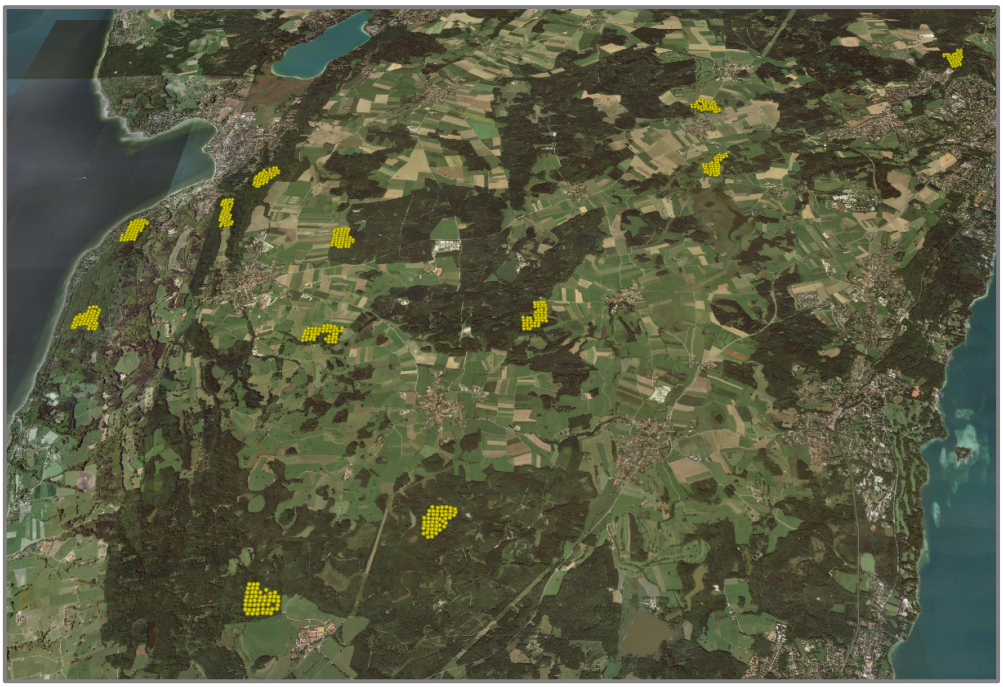

Figure 4. *Study sites with the geographic position of each of the nest boxes depicted in yellow circles. Nest boxes were overlaid on a Google Earth image.*

References

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

Bergmüller, R. & Taborsky, M. 2010. Animal personality due to social niche specialisation. *Trends in Ecology & Evolution*, **25**, 504–11.

Biro, P. A. & Adriaenssens, B. 2013. Predictability as a Personality Trait: Consistent Differences in Intraindividual Behavioral Variation. *The American Naturalist*, **182**, 621–629.

Both, C. 1998. Density dependence of clutch size: habitat heterogeneity or individual adjustment? *Journal of Animal Ecology*, **67**, 659–666.

Brommer, J. E. & Rattiste, K. 2008. “Hidden” reproductive conflict between mates in a wild bird population. *Evolution; international journal of organic evolution*, **62**, 2326–2333.

Brommer, J. E., Korsten, P., Bouwman, K. M., Berg, M. L. & Komdeur, J. 2007. Is extrapair mating random? On the probability distribution of extrapair young in avian broods. *Behavioral Ecology*, **18**, 895–904.

Browne, W. & McCleery, R. 2007. Using cross-classified multivariate mixed response models with application to life history traits in great tits (Parus major). *Statistical Modelling*, **7**,

217–238.

Cleasby, I. R., Nakagawa, S. & Schielzeth, H. 2014. Quantifying the predictability of behaviour: statistical approaches for the study of between-individual variation in the within-individual variance. *Methods in Ecology and Evolution*, n/a–n/a.

Dall, S. R. X., Houston, A. I. & McNamara, J. M. 2004. The behavioural ecology of personality: consistent individual differences from an adaptive perspective. *Ecology Letters*, **7**, 734–739.

Darwin, C. 1859. *The origin of species by means of natural selection.* London: Jhon Murray.

Dingemanse, N. J. & Wolf, M. 2010. Recent models for adaptive personality differences: a review. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **365**, 3947–58.

Dingemanse, N. J. & Wolf, M. 2013. Between-individual differences in behavioural plasticity within populations: causes and consequences. *Animal Behaviour*, **85**, 1–9.

Dingemanse, N. J., Kazem, A. J. N., Réale, D. & Wright, J. 2010. Behavioural reaction norms: animal personality meets individual plasticity. *Trends in Ecology & Evolution*, **25**, 81–9.

Duckworth, R. A. 2008. The role of behavior in evolution: a search for mechanism. *Evolutionary Ecology*, **23**, 513–531.

Enquist, M. & Leimar, O. 1983. Evolution of fighting behaviour: Decision rules and assessment of relative strength. *Journal of Theoretical Biology*, **102**, 387–410.

Falconer, D. S. & Mackay, T. F. C. 1996. *Introduction to Quantitative Genetics.* Longman.

Gross, M. 1996. Alternative reproductive strategies and tactics: diversity within sexes. *Trends in Ecology & Evolution*, **11**, 92–97.

Hamilton, W. D. 1964. The genetical evolution of social behaviour. I. *Journal of Theoretical Biology*, **7**, 1–16.

Huey, R. B., Hertz, P. E. & Sinervo, B. 2003. Behavioral drive versus behavioral inertia in evolution: a null model approach. *The American Naturalist*, **161**, 357–66.

Kokko, H. 2005. Mate guarding, male attractiveness, and paternity under social monogamy. *Behavioral Ecology*, **16**, 724–731.

Kölliker, M., Brinkhof, M. W., Heeb, P., Fitze, P. S. & Richner, H. 2000. The quantitative genetic basis of offspring solicitation and parental response in a passerine bird with biparental care. *Proceedings. Biological sciences / The Royal Society*, **267**, 2127–32.

Krebs, J. R. 1982. Territorial defence in the great tit : Do residents always win? *Behavioral*

Ecology and Sociobiology, **11**, 185–194.

Krebs, J. R. & Davies, N. B. 1997. *Behavioural Ecology: An Evolutionary Approach.* Blackwell Science.

Kvarnemo, C. & Simmons, L. 2013. Polyandry as a mediator of sexual selection before and after mating. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **368**, 20120042.

Laskowski, K. L. & Pruitt, J. N. 2014. Evidence of social niche construction: persistent and repeated social interactions generate stronger personalities in a social spider. *Proceedings. Biological sciences / The Royal Society*, **281**, 20133166.

Lynch, M. & Walsh, B. 1998. *Genetics and analysis of quantitative traits.* Sunderland, Mass.: Sinauer.

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

Maynard-Smith, J. & Price, G. R. 1973. The Logic of Animal Conflict. *Nature*, **246**, 15–18.

McGlothlin, J. W. & Brodie, E. D. 2009. How to measure indirect genetic effects: the congruence of trait-based and variance-partitioning approaches. *Evolution; international journal of organic evolution*, **63**, 1785–1795.

McGlothlin, J. W., Moore, A. J., Wolf, J. B. & Brodie, E. D. 2010. Interacting phenotypes and the evolutionary process. III. Social evolution. *Evolution; international journal of organic evolution*, **64**, 2558–2574.

McNamara, J. M., Stephens, P. a, Dall, S. R. X. & Houston, A. I. 2009. Evolution of trust and trustworthiness: social awareness favours personality differences. *Proceedings. Biological Sciences / The Royal Society*, **276**, 605–13.

Merilä, J., Sheldon, B. & Kruuk, L. 2001. Explaining stasis: microevolutionary studies in natural populations. *Genetica*, **112-113**, 199–222.

Montiglio, P., Ferrari, C. & Réale, D. 2013. Social niche specialization under constraints: personality, social interactions and environmental heterogeneity. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **368**, 20120343.

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

Morrissey, M. B. 2014. Selection and Evolution of Causally Covarying Traits. *Evolution;*

international journal of organic evolution, 1748–1761.

Nussey, D. H., Wilson, a J. & Brommer, J. E. 2007. The evolutionary ecology of individual phenotypic plasticity in wild populations. *Journal of Evolutionary Biology*, **20**, 831–44.

Patrick, S. C., Chapman, J. R., Dugdale, H. L., Quinn, J. L. & Sheldon, B. C. 2012. Promiscuity, paternity and personality in the great tit. *Proceedings. Biological sciences / The Royal Society*, **279**, 1724–1730.

Réale, D., Reader, S. M., Sol, D., McDougall, P. T. & Dingemanse, N. J. 2007. Integrating animal temperament within ecology and evolution. *Biological reviews of the Cambridge Philosophical Society*, **82**, 291–318.

Roff, D. A. 1992. *Evolution Of Life Histories: Theory and Analysis.* Springer.

Schlichting, C. & Pigliucci, M. 1998. *Phenotypic Evolution: A Reaction Norm Perspective.* Sinauer.

Sheldon, B. C., Kruuk, L. E. B. & Merila, J. 2003. When environmental variation short-circuits natural selection. **18**, 207–209.

Stearns, S. 1989. Trade-offs in life-history evolution. *Functional ecology*, **3**, 259–268.

Svensson, L. 1992. *Identification guide to European passerines.* L. Svensson.

Van de Pol, M. 2012. Quantifying individual variation in reaction norms: how study design affects the accuracy, precision and power of random regression models. *Methods in Ecology and Evolution*, **3**, 268–280.

Van Oers, K., Drent, P. J., Dingemanse, N. J. & Kempenaers, B. 2008. Personality is associated with extrapair paternity in great tits, Parus major. *Animal Behaviour*, **76**, 555–563.

Wagner, G. P. 2001. *The Character Concept in Evolutionary Biology.* San Diego, CA: Academic Press-Elsevier Science.

Webster, M., Pruett-Jones, S., Westneat, D. & Arnold, S. 1995. Measuring the effects of pairing success, extra-pair copulations and mate quality on the opportunity for sexual selection. *Evolution; international journal of organic evolution*, **49**, 1147–1157.

West-eberhard, M. J. 1979. Sexual Selection, Social Competition, and Evolution. *Proceedings of the American Philosophical Society*, **123**, 222–234.

Westneat, D. F. 2012. Evolution in response to social selection: the importance of interactive effects of traits on fitness. *Evolution; international journal of organic evolution*, **66**, 890–5.

Westneat, D. & Fox, C. W. 2010. *Evolutionary behavioral ecology.* New York: Oxford University Press.

Westneat, D. F. & Stewart, I. R. K. 2003. Extra-pair Paternity in Birds: Causes, Correlates, and Conflict. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 365–396.

Westneat, D. F., Wright, J. & Dingemanse, N. J. 2014. The biology hidden inside residual within-individual phenotypic variation. *Biological Reviews*, n/a–n/a.

Wilson, A. J. 2013. Competition as a source of constraint on life history evolution in natural populations. *Heredity*, 1–9.

Wilson, A. J., Morrissey, M. B., Adams, M. J., Walling, C. a, Guinness, F. E., Pemberton, J. M., Clutton-Brock, T. H. & Kruuk, L. E. B. 2011. Indirect genetics effects and evolutionary constraint: an analysis of social dominance in red deer, Cervus elaphus. *Journal of Evolutionary Biology*, **24**, 772–83.

Wolf, J. B. 2003. Genetic architecture and evolutionary constraint when the environment contains genes. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 4655–4660.

Wolf, M., van Doorn, G. S., Leimar, O. & Weissing, F. J. 2007. Life-history trade-offs favour the evolution of animal personalities. *Nature*, **447**, 581–4.

Chapter 1

Characterizing behavioral ‘characters': an evolutionary framework

Yimen G. Araya-Ajoy & Niels J. Dingemanse

Proceedings of the Royal Society B (2014), 282, 20132645

PROCEEDINGS OF THE ROYAL SOCIETY B

rspb.royalsocietypublishing.org

Research

Cite this article: Araya-Ajoy YG, Dingemanse NJ. 2014 Characterizing behavioural 'characters': an evolutionary framework. *Proc. R. Soc. B* **281**: 20132645.
http://dx.doi.org/10.1098/rspb.2013.2645

Received: 9 October 2013
Accepted: 19 November 2013

Subject Areas:
behaviour, evolution, ecology

Keywords:
modularity, behavioural characters, personality, psychology, plasticity, variance components

Author for correspondence:
Yimen G. Araya-Ajoy
e-mail: yimencr@gmail.com

Electronic supplementary material is available at http://dx.doi.org/10.1098/rspb.2013.2645 or via http://rspb.royalsocietypublishing.org.

Royal Society Publishing

Characterizing behavioural 'characters': an evolutionary framework

Yimen G. Araya-Ajoy[1] and Niels J. Dingemanse[1,2]

[1]Evolutionary Ecology of Variation Group, Max Planck Institute for Ornithology, Seewiesen, Germany
[2]Behavioural Ecology, Department Biology II, Ludwig-Maximilians-University of Munich, Planegg-Martinsried, Germany

Biologists often study phenotypic evolution assuming that phenotypes consist of a set of quasi-independent units that have been shaped by selection to accomplish a particular function. In the evolutionary literature, such quasi-independent functional units are called 'evolutionary characters', and a framework based on evolutionary principles has been developed to characterize them. This framework mainly focuses on 'fixed' characters, i.e. those that vary exclusively between individuals. In this paper, we introduce multi-level variation and thereby expand the framework to labile characters, focusing on behaviour as a worked example. We first propose a concept of 'behavioural characters' based on the original evolutionary character concept. We then detail how integration of variation between individuals (cf. 'personality') and within individuals (cf. 'individual plasticity') into the framework gives rise to a whole suite of novel testable predictions about the evolutionary character concept. We further propose a corresponding statistical methodology to test whether observed behaviours should be considered expressions of a hypothesized evolutionary character. We illustrate the application of our framework by characterizing the behavioural character 'aggressiveness' in wild great tits, *Parus major*.

1. Introduction

Biologists often study phenotypic evolution assuming that phenotypes consist of a set of quasi-independent units or parts that have been shaped by selection to accomplish a particular function [1,2]. Consequently, the success of evolutionary research programmes depends to a large degree on whether such functional units have been properly characterized. For this reason, evolutionary biologists have developed an appealing conceptual framework (detailed below), in which these functional units are called 'evolutionary characters' [3]. Notably, despite the importance of labile characters in mediating interactions between organisms and their environment [4], they have not been fully integrated into this framework. This is in part because labile characters (e.g. behaviour) vary both between and within individuals; previous implementations have instead primarily focused on fixed phenotypes (e.g. structural size). In this paper, we expanded this framework to integrate (any) multi-level structure and illustrate its application by characterizing behavioural phenotypes. We introduce a definition of 'behavioural characters' and propose a general methodology that enables empirical testing of novel hypotheses concerning the question of whether observed behaviours can be considered expressions of a hypothesized evolutionary character.

Central to our framework is the concept of evolutionary characters, which can be defined as parts of an organism that exhibit causal coherence in their expression and play a causal role in a biological process [3]. This definition has two important characteristics. First, its causal coherence refers to a set of inter-related mechanisms that are involved in the character's expression and makes it quasi-independent from other characters [5]. This 'modularity' is what enables the character to respond adaptively to selection [6]. Second, its explicit link to a biological process implies that a character is a 'functional unit' used by an organism for a particular task. An evolutionary phenotypic

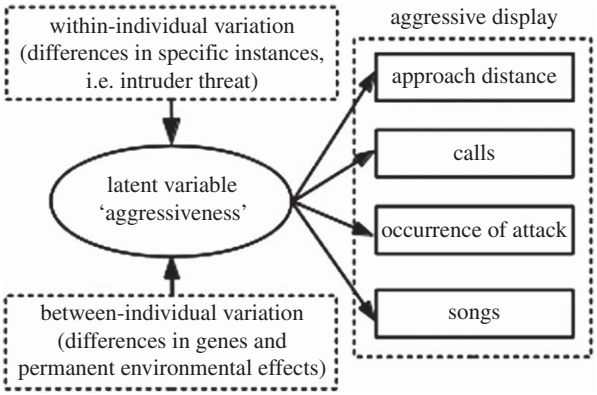

Figure 1. Diagram of the multi-level evolutionary character concept applied to avian agonistic behaviour. The behavioural character 'aggressiveness' is represented as a latent variable affecting the expression of observed behaviours (calls, approach distance, occurrence of attack and songs). The hypothesized expression of the latent variable is plastic within the same individual, as it varies as a function of environmental conditions (top-left), but also differs between individuals owing to genetic and environmental effects specific to the individual (lower-left). Consequently, expressed behaviours are correlated in a similar fashion between versus within individuals.

module or 'character' is thus composed of several elements that are functionally related [7]. Characters are themselves in turn hierarchically structured, where a functional unit can be considered a part of a higher level unit [8]. For example, one can consider the human hand as a character that is composed of five fingers and is used to grab objects and use tools. Each finger has a specific function and can be considered a character by its own, but because all fingers need to be used as a coherent functional unit when using tools, they must be tightly correlated in terms of length, shape and neurological underpinning [9]; therefore, fingers of the same hand respond as a unit to selective forces and can be viewed as expressions of the same character. We propose to apply this general logic to behaviour and define behavioural characters by the causal coherence underlying their expression and the function that they accomplish for the organism.

We illustrate our behavioural character concept using aggressiveness displayed by territorial male great tits, *Parus major* (figure 1). We view aggressiveness as a behavioural character that dictates how an organism responds to agonistic interactions. Great tits express a wide array of behaviours during such encounters [10] that jointly execute a specific function: displacing intruders. We therefore *a priori* visualize aggressiveness as an unobserved—statistically called 'latent'—variable that affects multiple behaviours used in aggressive displays (visualized in figure 1 by arrows connecting the latent variable with the expressed behaviours). For example, during highly aggressive interactions, male great tits respond to a conspecific intrusion by calling while approaching and attacking if the intruder does not withdraw. By contrast, during less aggressive interactions, males sing from far away rather than calling and approaching close. Proximately, this functional coherence is owing to common mechanisms affecting the expression of all behaviours of the display (i.e. through pleiotropic effects of genetic or environmental factors; [11]). This common (neurological or physiological) pathway enables different behaviours to be expressed as a functional unit. It is this proximate mechanism that evolves in response to selection and that represents the character [12]. We note that the terms phenotypic 'character' versus 'trait' are used interchangeably in the evolutionary literature. Traits are sometimes defined directly as observable variables that are biologically relevant; here, we simply call measured quantities 'observable variables' and refer to 'characters' as the inferred theoretical entities underlying the expression of functionally related observable variables. This borrows from the statistical and psychological literature where a distinction is made between attributes that are directly measurable versus those reflecting underlying unobservable quantities [13,14]. We thus propose that behavioural characters represent unmeasured 'latent' variables that can be inferred from the expression patterns of behavioural observables.

Other fields of biology, especially human personality research, have a long and productive history of studying behaviour using latent variables [14]. Our approach is distinctly different because we explicitly address the issue of how one might integrate behavioural variation between individuals (cf. 'animal personality' [15]) and variation within individuals (cf. adaptive 'individual plasticity' [16]) when studying these latent variables (behavioural characters) from an evolutionary perspective (detailed further in the Discussion). We will continue our worked example to explain this unique aspect of our approach. If there is a latent variable (aggressiveness) affecting the expression of the different agonistic behaviours, it will cause between-individual and within-individual correlations between the agonistic behaviours (cf. [17]). On the one hand, between-individual differences in aggressiveness owing to genetic differences or early-life experiences (visualized in figure 1 by the lower dashed box with arrows pointing to the latent variable) will result in between-individual correlations among behaviours of the display. Aggressive individuals should, for example, on average have high values for call rate as well as higher tendency to approach intruders. On the other hand, within-individual plastic responses to environmental changes should result in correlated changes in all behaviours of the display within the same individual (visualized in figure 1 by the upper dashed box with arrows pointing to the latent variable), resulting in within-individual correlations. If an individual increases its level of aggressiveness, its call rate should increase and it should approach the intruder closer. Decomposition and comparison of behavioural correlations within versus between individuals therefore provides clues about whether a common underlying mechanism might underpin behavioural variation at different levels.

The behavioural character concept consequently comes with predictions about patterns of (co)variation between behavioural expressions of a character. First, each of the observed behaviours should show between-individual variation (i.e. non-zero repeatability) and part of this variation should be owing to individual differences in a latent variable, provided that the population harbours between-individual variation in the behavioural character. Second, behavioural expressions of the character should change in concert within the same individual in response to environmental change ('integration of plasticity'; [18]), provided that the behavioural character is plastic within individuals. Third, similar non-zero behavioural correlations are expected between versus within individuals provided that the character also varies at both levels. Fourth, correlations between expressed behavioural observables should be the same in different environments in which the character is expressed

(e.g. breeding versus non-breeding contexts), and significant cross-environment correlations should exist if the same mechanism (character) affected the expression of behavioural observables in different environments [11]. Finally, a character should be quasi-independent from other characters to respond to selection as a unit [6]. Behavioural expressions of a character should therefore show some degree of independence from other behavioural characters.

We illustrate our thesis by analysing four behaviours that great tit males use when confronted with a territorial intrusion. We tested the hypothesis that these four behavioural observables were expressions of the behavioural character 'aggressiveness'. To do so, each male was subjected to a 'standardized territorial intrusion' four times per breeding season (year): twice during the egg-laying period of its social mate, at which time intrusions should increase risk of paternity loss [19] and consequently elicit a relatively aggressive response and twice when its social mate was incubating the clutch, at which time intrusions should not increase perceived risk of paternity loss and consequently elicit less of an aggressive response. We tested whether the data supported the hypothesis that the behavioural observables were indeed expressions of the same character ('aggressiveness').

We performed a four-step data analysis: we first ran univariate analyses where, for each of the behavioural observables separately, we estimated the amount of variance between and within individuals, as well as the level of behavioural plasticity with respect to breeding context. We expected that all behavioural observables would have a repeatable component and show a plastic response to the relative perceived threat of the intruder, provided that they represented expressions of the same repeatable but plastic behavioural character. Therefore as the second step, we quantified correlations between the different behavioural observables, asking whether they were correlated as hypothesized at each hierarchical level (i.e. between and within individuals) and in each environment (i.e. during laying and incubation). The integration of behavioural observables across environments was investigated using a character state approach [20] and assessed by testing whether correlations within environments (breeding contexts) and across environments ('cross-environment' correlations; [20]) were consistent with the presence of a single common underlying mechanism. As the third step, we statistically evaluated the amount of support for the presence of a context-general latent variable. Finally, we asked whether 'aggressiveness' constituted a quasi-independent module by evaluating whether it was distinct from other presumed aspects of risk-taking behaviour, for example level of activity in a novel environment.

2. Material and methods

(a) Experimental protocol

We studied 12 nest box populations of great tits in southern Germany (for details, see electronic supplementary material, appendix S1). Simulated territorial intrusions (i.e. aggression tests) were performed in the breeding seasons of 2010–2012. A taxidermic mount of a male great tit was presented as a visual stimulus with a playback song as an acoustic stimulus (detailed below). In each year, each male was subjected to four aggression tests during its first breeding attempt (defined as attempts initiated within 30 days after the first egg of the year in all of the plots was found; [21]). Each male was subjected to two simulated territorial intrusions during egg-laying (1 and 3 days after its first egg was observed) and two during incubation (1 and 3 days after clutch incubation was confirmed). Owing to logistical constraints, the interval between first and repeat trials within-breeding context was more than 2 days for 7% of the 1150 repeat tests.

Aggression tests were conducted between 7.00 and 12.00; the specific time was semirandomly assigned. The taxidermic mount was presented 1 m away from the subject's nest-box on a 1.2 m wooden pole protected by a green wire mesh (see the electronic supplementary material, figure S1). Fifteen mounts and 14 playback song stimuli (recorded from German and Dutch populations) were constructed, enabling us to test whether the assayed behaviours represented responses to great tit mounts and songs in general rather than responses to their specific characteristics [22]; one mount and one song (broadcasted with a Samsung U5 Digital Audio Player connected to a Radioshack Mini Amplifier) were randomly allocated to each test. One of 25 observers performed the observation at a distance of 15 m.

Following the onset of a focal test, we recorded the behaviour of the focal male for a period of 3 min after it had entered a 15 m radius around the box. The observer counted the number of calls and songs, estimated the minimum distance to the mount ('approach distance') and noted whether the subject attacked the mount (jumping on the wire mesh of the mount; 'occurrence of attack'). (Descriptive statistics of each observable (cf. mean, range and standard deviation) are given in the electronic supplementary material, table S1.) For ease of interpretation, approach distance was multiplied by −1 (i.e. higher values represented a more aggressive response) in all the statistical analyses. Subjects that did not arrive within 15 min were scored as non-responsive. We performed 657 tests in 2010, 652 in 2011 and 937 in 2012, reflecting yearly breeding densities. Male identity was known for 1593 tests; in 1285 (80%) of these tests, the male responded. Analyses were based on these 1285 aggression tests, representing 365 unique (i.e. ringed) males. The number of responsive tests varied between males depending on number of years present and number of responses: 10 tests (n males = 1), 9 ($n = 3$), 8 ($n = 11$), 7 ($n = 17$), 6 ($n = 22$), 5 ($n = 21$), 4 ($n = 80$), 3 ($n = 104$) 2 ($n = 66$), 1 ($n = 40$).

(b) Statistical analyses

(i) Univariate mixed-effect models

We modelled variation in each of the agonistic behaviours separately as a function of (fixed effects) breeding context (laying versus incubation), test sequence within-breeding context (first versus second trial), year (2010, 2011, 2012) and time of the day (measured as minutes after sunrise and expressed as the deviation from the average time of all tests). Random intercepts were included for the identity of the observer ($n = 25$ levels), population ($n = 12$), playback song ($n = 14$), taxidermic mount ($n = 15$) and subject male ($n = 365$). We used the following error structure: approach distance was square root transformed and modelled with Gaussian errors, number of songs and calls (untransformed) modelled with Poisson errors and occurrence of attack (yes/no) with binomial errors. Adjusted repeatabilities were subsequently calculated as the between-individual variance divided by the sum of the between-individual and the residual variance [23].

(ii) Multi-variate mixed-effect models

Between- and within-individual correlations were estimated by fitting the assayed behaviours (approach distance, calls, songs and occurrence of attack) as four response variables into a single multi-variate mixed-effect model with random intercepts for individual identity. Further fixed or random effects were not included because our univariate analyses revealed that their effects were of minor importance (see Results and table 1). Breeding context

Table 1. Sources of variation in four agonistic behaviours based on simulated territorial intrusion experiments applied to great tits in southern Germany. (Estimates were derived, separately for each agonistic behaviour, from univariate mixed-effect models with random intercepts for individual (1–365), population (1–12), observer (1–25), taxidermic model (1–15) and playback song identity (1–14). Breeding context (laying versus incubation), test sequence within-breeding context (first versus second), time of day and year (2010, 2011, 2012) were fitted as fixed effects ($n = 1285$ tests). We give point estimates for each fixed (β; mean) and random (σ^2; variance) parameter, as well as adjusted repeatabilities, with their 95% CI.)

	calls	approach distance[a]	occurrence of attack	songs
fixed effects	β (95% CI)	β (95% CI)	β (95% CI)	β (95% CI)
intercept[b]	−1.05 (−2.00, −0.38)	−2.49 (−2.65, −2.14)	−1.89 (−3.54, −0.56)	1.92 (1.68, 2.16)
breeding context	−1.96 (−2.36, −1.44)	−0.61 (−0.71, −0.47)	−2.37 (−3.04, −1.49)	0.47 (0.34, 0.59)
sequence	−0.43 (−0.94, −0.07)	−0.10 (−0.21, 0.03)	−0.33 (−0.85, 0.26)	0.11 (0.00, 0.25)
time of day	−0.23 (−0.51, −0.02)	0.07 (−0.01, 0.12)	0.00 (−0.01, 0.00)	−0.01 (−0.08, 0.05)
year 2011	1.06 (0.39, 1.88)	0.11 (−0.15, 0.30)	0.69 (−0.35, 1.44)	−0.11 (−0.29, 0.16)
year 2012	1.10 (0.33, 1.78)	−0.05 (−0.33, −0.16)	0.02 (−1.28, 0.59)	−0.21 (−0.41, 0.06)
random effects	σ^2 (95% CI)	σ^2 (95% CI)	σ^2 (95% CI)	σ^2 (95% CI)
individual	3.40 (2.13, 4.95)	0.40 (0.32, 0.56)	2.20 (0.01, 5.43)	0.40 (0.26, 0.51)
population	0.01 (0.00, 1.52)	0.00 (0.00, 0.08)	0.01 (0.00, 0.48)	0.00 (0.00, 0.09)
observer	0.00 (0.00, 0.28)	0.04 (0.00, 0.09)	0.00 (0.00, 0.32)	0.01 (0.00, 0.07)
model	0.00 (0.00, 0.37)	0.00 (0.00, 0.03)	0.00 (0.00, 0.67)	0.00 (0.00, 0.02)
song	0.06 (0.00, 0.12)	0.00 (0.00, 0.01)	0.00 (0.00, 0.17)	0.00 (0.00, 0.04)
residual[c]	10.19 (8.55, 12.06)	1.10 (1.00, 1.20)	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)
repeatability	r (95% CI)	r (95% CI)	r (95% CI)	r (95% CI)
	0.21 (0.14, 0.29)	0.29 (0.22, 0.35)	0.38 (0.13, 0.61)	0.25 (0.20, 0.32)

[a]Approach distance was multiplied by −1 prior to analysis.
[b]Reference categories for fixed effects were set to 'laying' (breeding context), '1st' (sequence), 2010 (year) and population mean time of the day.
[c]Residual error distributions were binomial (occurrence of attack), Gaussian (approach distance) or Poisson (calls, songs).

strongly affected all of the behaviours (table 1) but was not included in the model because we wanted the within-individual covariance matrix to capture all sources of within-individual plasticity. Behaviour-specific error structure was applied as detailed above. Notably, the within-individual variance of 'occurrence of attack' was fixed to one because it is not estimable for binary data [13]; within-individual correlations with this variable should consequently be treated with caution. Exclusion of this response variable did not change our general findings (see electronic supplementary material, table S2*b*).

Within- and cross-breeding context correlations were estimated at the between-individual level by treating each of the four behavioural observables as a distinct response variable for each breeding context (e.g. 'songs during laying' and 'songs during incubation'), resulting in a multi-variate mixed-effect model with eight response variables and random intercepts for individual identity. We consequently estimated, within the same model, between-individual correlations within *and* across breeding contexts. This model estimated 28 between-individual correlations (six within-context correlations among all four behavioural observables × 2 contexts + 16 across-context correlations). Within-individual cross-context covariances were non-estimable (because the two breeding contexts cannot be experienced at the same time) and were therefore constrained to zero [17]; further fixed or random effects were not included (detailed above).

To assess whether the behaviours were correlated as expected according to the behavioural character concept, we compared the similarity between the posterior distributions (defined below) of pairwise correlations between versus within individuals and between laying versus incubating, using the 'overlapping coefficient' [24]. We further applied Mantel tests to assess whether the two matrices differed in correlation structure.

(iii) Structural equation modelling

We applied structural equation modelling (a statistical technique that includes confirmatory factor analysis as a special case) to the between-individual covariance matrix derived from the mixed-effect model with eight response variables (detailed above). We evaluated relative support for each of four *a priori* considered scenarios (based upon their relative AIC-values): (i) the absence of any latent variable (figure 2*a*); (ii) the presence of a single latent variable affecting all behaviours in both contexts (figure 2*b*); (iii) the presence of two context-specific latent variables (figure 2*c*) and (iv) the presence of two correlated but context-specific latent variables (figure 2*d*).

(iv) Quasi-independence of behavioural modules

We tested for quasi-independence of the hypothesized aggressiveness module by assessing whether the four agonistic behaviours (occurrence of attack, approach distance, calls and songs) were correlated with another observed behaviour, the individual's level of activity when placed into a novel environment (see [25] and electronic supplementary material, appendix S2). We estimated the correlations between the four hypothesized behavioural expressions of the character aggressiveness and activity in a novel environment by fitting them all as response variables into a multivariate mixed-effect model with random intercepts for individual identity (1–277).

(v) Parameter estimation methods

We used R statistical environment v. 3.0.2 for all statistical analyses [26]. Mixed-effect models were fitted using Monte Carlo Markov chains in the MCMCglmm package [27], which retrieves posterior distributions of estimated parameters. We subsequently

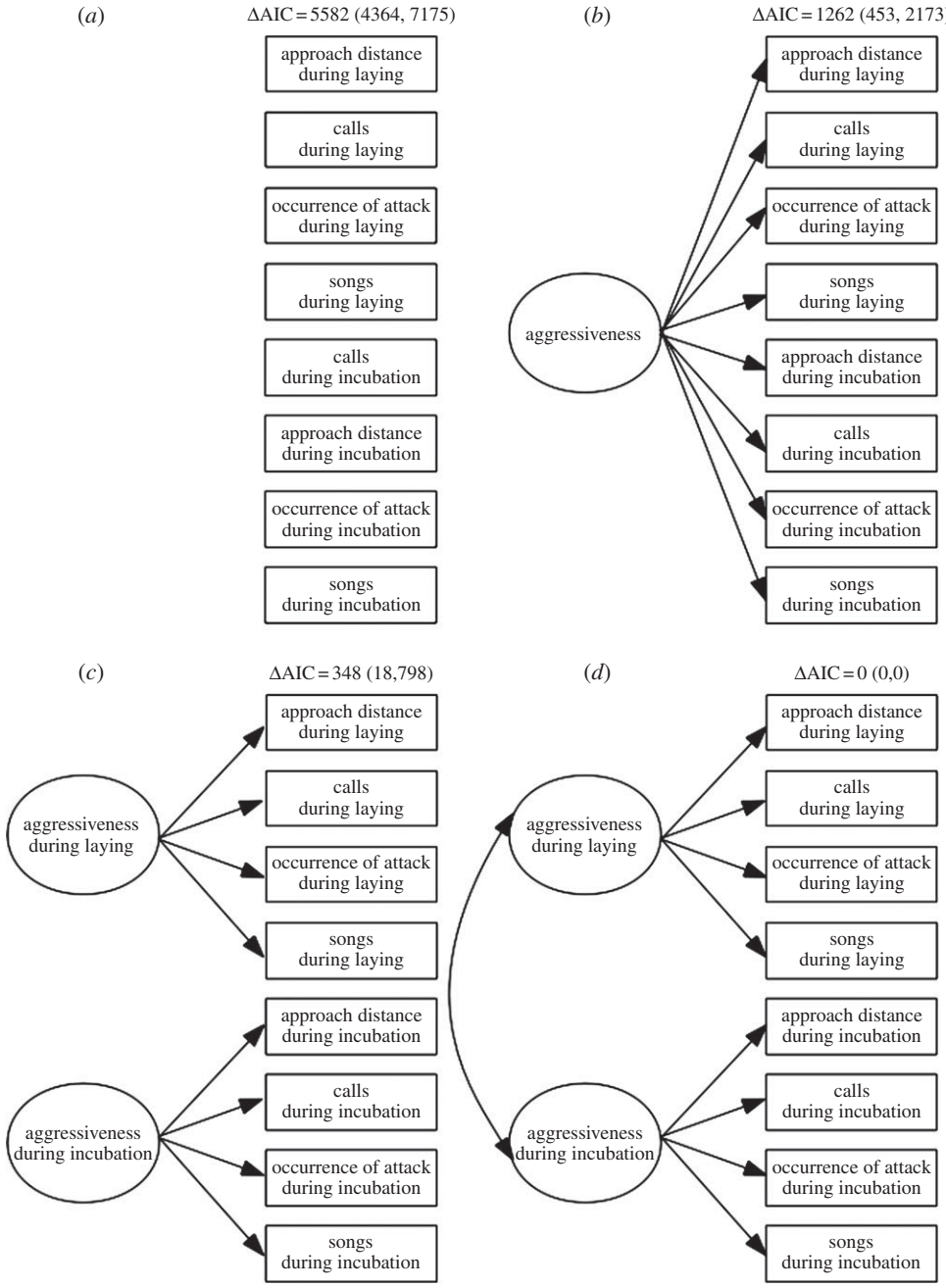

Figure 2. Four models (hypotheses) explaining covariance structure among agonistic behaviours assayed during laying and incubation in wild great tits. Model (*a*) proposes a scenario where each combination of observables and breeding stage is underpinned by a separate factor (the null model); model (*b*) hypothesizes a common factor ('module') underpinning all observables regardless of breeding context, whereas model (*c*) hypothesizes a separate module for each breeding context; model (*d*) expands upon this scenario by hypothesizing that those modules are themselves submodules influenced by a common factor.

calculated the mode and 95% credible interval (CI) for each parameter. This Bayesian approach allows for uncertainty to be appropriately carried forward to follow-up analyses [28]. Structural equation models were fitted with the 'sem' package [29]. Model implementation and procedures used for taking forward uncertainty from one analysis to the next are detailed in the electronic supplementary material, appendix S3.

3. Results

(a) Sources of variation in behavioural observables

A substantial part of phenotypic variation in each of the observed agonistic behaviours was explained by differences between individuals. CIs for repeatability were never close to zero, implying strong support for the presence of between-individual variation. Adjusted repeatability ranged between 0.21 and 0.38 (table 1). All behavioural observables changed with breeding context (table 1): individuals produced more calls, sang less, approached closer and were more likely to attack during laying compared with incubation (see electronic supplementary material, table S1). Effects of time of day, test sequence or year were not supported, except for calls that differed among years and decreased with time of day and sequence within-breeding context (table 1). The identity of the observer, mount or playback song explained little variation, if any at all (table 1).

(b) Between- versus within-individual correlations

Our mixed-effect model with four response variables (see Material and methods) provided strong support for non-zero correlations among all behavioural observables at the

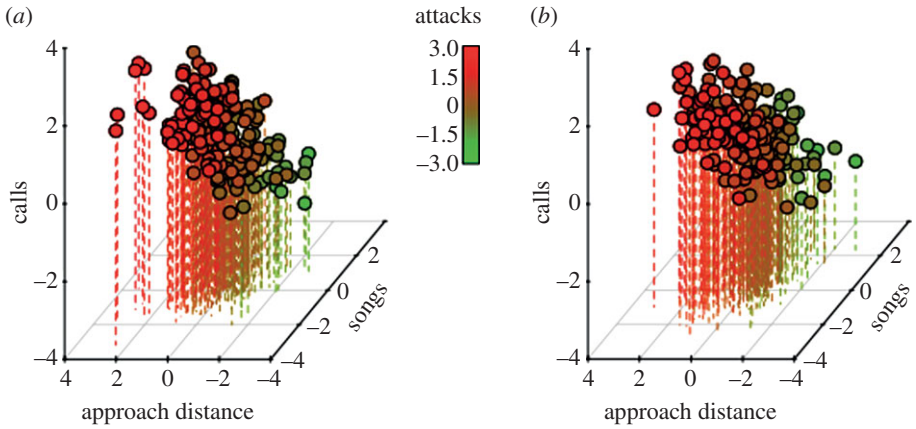

Figure 3. (*a*) Visual representation of between-individual correlations (plotted are the individuals' average values as deviations from the population mean value) and (*b*) within-individual correlations (plotted are the observations represented as deviations from individual mean values) between agonistic behaviours observed in wild great tits in Bavaria (Germany). Values were simulated from the correlation matrix estimated from a multi-variate mixed-effect model with random intercepts for individual identity (1–365) and calls, approach distance, occurrence of attack and songs fitted as response variables. We plot here predicted values on their untransformed latent scale.

between-individual level (figure 3*a*; electronic supplementary material, table S2). Individuals that on average (across all observations) approached the mount relatively closely also called at relatively high rates, produced fewer songs and were more likely to attack the model compared with individuals that on average did not approach closely. Within-individual correlations showed the same pattern (figure 3*b*; electronic supplementary material, table S2): during observations where an individual approached the dummy relatively closely, it would also call relatively much but sing relatively little compared with observations of the same individual where it approached less closely. These findings imply that the assayed agonistic behaviours changed in concert as hypothesized. Posterior distributions of pairwise correlations within versus between individuals overlapped substantially (see electronic supplementary material, table S2), providing strong support for the hypothesis that behavioural correlations did not differ between levels. This was confirmed by matrix-wide statistical comparisons (Mantel test: r (95% CI) = 0.88 (0.76–0.96)). Taken together, these findings support the hypothesis that the same latent variable (character) affected the expression of the agonistic behavioural observables within versus between individuals.

(c) Between-individual correlations within- versus across-breeding contexts

Signs and magnitudes of between-individual within-breeding context correlations were very similar for the two breeding contexts (table 2*b*): posterior distributions overlapped considerably (see electronic supplementary material, table S3). For example, the correlation between calls and approach distance was (point estimate (95% CI)) 0.67 (0.53, 0.77) during laying and 0.53 (0.35, 0.67) during incubation (overlap: 0.32). This similarity was confirmed by matrix-wide statistical comparisons (Mantel test: r (95% CI) = 0.98 (0.91–0.99)).

Most behavioural observables showed 'significant' positive between-individual cross-breeding context correlations (i.e. most CIs did not overlap zero; table 2*a*). In other words, individuals that had relatively high average values during laying also had relatively high average values during incubation, suggesting that the same behavioural observable was proximately underpinned by the same mechanism when expressed in different contexts. Upper CI nevertheless never included 1.00 (calls: 0.72; approach distance: 0.60; occurrence of attack: 0.65; songs: 0.62), implying that their between-individual variances were also shaped—though only partly—by context-specific proximate factors [19]. Crossbreeding context correlations between different behavioural observables were of the same sign as their within-context counterparts, but the former correlations were less strong (table 2*b*), again suggesting some level of context-specific expression of between-individual variance (i.e. hierarchical structure) in the presumed behavioural character.

(d) Structural equation modelling

Cross-context correlations were substantial but their within-context counterparts were tighter (table 2), implying context-specific but correlated submodules affecting the expression of the behavioural observables (cf. model (*d*) in figure 2). Our comparison of four *a priori* considered structural equation models supported this interpretation: model (*d*) was the single best-supported model; the upper 95% CI of its ΔAIC value did not overlap with the lower 95% CI of other models (figure 2). The presumed context-specific submodules (cf. latent variables) were, as expected, positively correlated (r (95% CI): 0.59 (0.24, 0.78); figure 4).

(e) Quasi-independence of behavioural modules

The four observed agonistic behaviours were, as expected, not associated with activity in a novel environment. There was very little statistical support for correlations between observables that were *a priori* hypothesized expressions of the character 'aggressiveness' and activity in a novel environment. (see electronic supplementary material, table S4).

4. Discussion

This paper proposed an approach for the inclusion of 'labile characters' into the evolutionary character framework [3] and introduced a corresponding statistical methodology to test whether labile observables can be considered expressions of a hypothesized evolutionary character. We used the labile behavioural character 'aggressiveness' in great tits as a worked example to show that the character concept has novel

Table 2. Between-individual correlations (*r*) between four agonistic behaviours within- and across-breeding contexts (laying versus incubation). (Estimates were derived from a cross-environment multi-variate mixed-effect model where each of four agonistic behaviours (calls, approach distance, occurrence of attack and songs) was fitted as a separate response variable for each breeding context (i.e. eight response variables), with random intercepts for individual identity. (*a*) Between-individual correlations between the same agonistic behaviour across the two contexts; (*b*) between-individual correlations between two different agonistic behaviours within breeding contexts (i.e. both behaviour 1 and 2 are measured within the same context) and across breeding contexts (i.e. behaviour 1 is measured during laying and behaviour 2 instead during incubation or *vice versa*). We give point estimates for each parameter with their 95% CI.)

	within-breeding context correlations		cross-breeding context correlations	
context 1 – context 2	laying – laying	incubation – incubation	laying – incubation	incubation – laying
behaviour 1 – behaviour 2	*r* (95% CI)	*r* (95% CI)	*r* (95% CI)	*r* (95% CI)
(*a*) same behaviour				
calls – calls	—	—	0.58 (0.34, 0.72)	
approach – approach	—	—	0.51 (0.31, 0.60)	
attack – attack	—	—	0.34 (−0.14, 0.65)	
songs – songs	—	—	0.45 (0.32, 0.62)	
(*b*) different behaviours				
calls – approach	0.67 (0.53, 0.77)	0.53 (0.35, 0.67)	0.32 (0.12, 0.49)	0.34 (0.12, 0.50)
calls – attack	0.72 (0.51, 0.83)	0.67 (0.34, 0.88)	0.28 (−0.16, 0.61)	0.27 (0.03, 0.50)
calls – songs	−0.80 (−0.88, −0.71)	−0.73 (−0.82, −0.62)	−0.37 (−0.54, −0.20)	−0.73 (−0.83, −0.62)
approach – attack	0.86 (0.78, 0.91)	0.64 (0.46, 0.78)	0.34 (−0.03, 0.63)	0.44 (0.22, 0.61)
approach – songs	−0.54 (−0.67, −0.39)	−0.28 (−0.44, −0.11)	−0.18 (−0.32, 0.03)	−0.34 (−0.48, −0.10)
attack – songs	−0.56 (−0.68, −0.40)	−0.44 (−0.66, −0.17)	−0.15 (−0.32, 0.08)	−0.31 (−0.60, 0.08)

predictions that are empirically testable when applied to multi-level phenotypes. Explicit to this framework is (i) that characters should be defined *a priori* as latent variables that affect functionally correlated observables, (ii) that if just one observable was measured, it would not be possible to validate whether it did reflect the character of interest, (iii) that both variation between and within individuals should explicitly be acknowledged and incorporated and (iv) that functionally unrelated observables also need to be measured to test for the quasi-independence of an hypothesized character from other ones.

(a) Novel predictions

Our multi-level implementation of the character concept introduced novel predictions that concern specific variance components [17] of observables. First, all labile observables that are *a priori* hypothesized expressions of a labile character should logically contain between-individual variance if the character itself contains between-individual variation. A statistical outcome where some but not all hypothesized expressions of a character showed between-individual variation would suggest that the hypothesis was false. Second, all labile observables should respond in concert to variation in the environment if they belong to a functional unit [11]. If one of the observables would not show a plastic response to a specific environmental gradient while others did, they would not all be expressions of the same character. Third, labile observables should correlate similarly at all hierarchical levels at which the latent variable varied. In summary, the characteristic multi-level nature of labile characters will enable researchers to test predictions that have not previously been considered in evolutionary character theory. We applied this logic to the between versus within individual level, but it would equally apply to others (e.g. between versus within populations; [30]). Assessment of similarity in between-individual correlation structure when comparing contexts (cf. laying versus incubation in our worked example) constitutes another test of the same idea.

(b) Empirical testing of predictions

In our worked example, we defined aggressiveness as a latent variable affecting the expression of behaviours used in aggressive interactions. We subsequently tested whether four behaviours used in aggressive interactions were indeed expressions of this labile character. We found between-individual differences in all assayed behaviours (table 1) that were partly attributable to hypothesized latent variables (figure 4). These observables were all plastic in a coordinated way as expected based upon level of intruder threat (effect of breeding context: table 1). This suggests a common underlying proximate mechanism that makes the aggressive display a functional unit. Patterns of correlation within and between individuals agreed with this interpretation: all expressed behaviours were associated, and in a very similar way, within and between individuals (figure 3). Sign and magnitude of the between-individual correlations also did not differ between breeding contexts (table 2), despite substantial levels of cross-context plasticity (table 1). Furthermore, an individual's typical value for a focal behaviour was repeatable across breeding contexts (i.e. cross-environment correlations were positive; table 2), supporting the notion of a common context-independent mechanism affecting all agonistic behaviours. At the same time, between-individual within-context correlations were somewhat tighter than their cross-context counterparts (table 2), implying partial context-specific modularity (figure 4). Those modules were positively correlated, implying that they represented submodules of an overarching context-independent latent

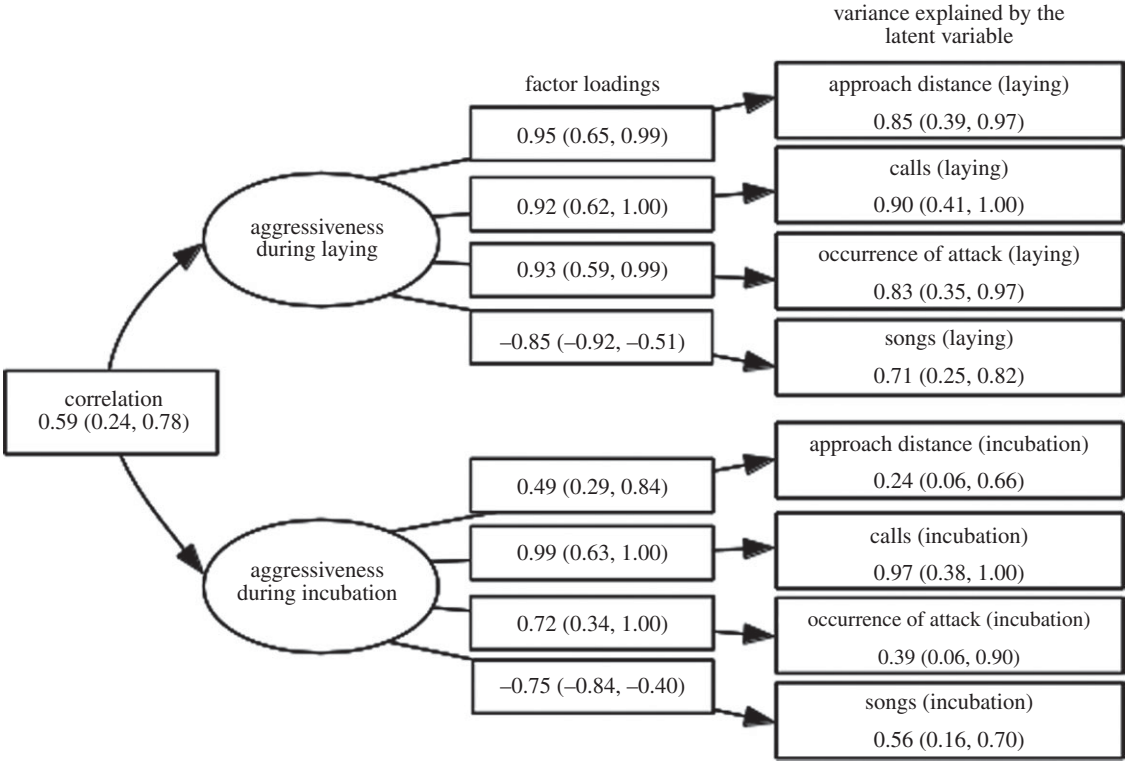

Figure 4. Parameter estimates of the structural equation model that best fitted our data.

variable affecting the aggressive display. This finding shows that the expression of the latent variable itself had a hierarchical structure, illustrating the level of detail about the structure of labile characters that can be derived by applying this framework. Furthermore, we showed that activity in a novel environment—a presumed observable expression of 'risk-taking behaviour' in non-social contexts [25]—was not significantly correlated to any of the agonistic behaviours (see electronic supplementary material, table S4), implying that the behavioural character 'aggressiveness' indeed represented a quasi-independent behavioural module at least with respect to this observed behaviour but potentially also from other risky behaviours in general.

(c) Why study behavioural characters?

The study of behaviour has a long history in fields of evolutionary biology (cf. animal behaviour and behavioural ecology), with research programmes focusing on a diverse array of topics, such as proximate causation, development and function of behaviour [31]. The proposed application of the evolutionary character concept in the study of behaviour will, in our opinion, greatly help researchers in deciding whether observed behaviours do or do not quantify the 'characters' that correspond to those for which adaptive theory has been developed. For example, theory predicts that between-individual variation in future fitness expectations can explain between-individual variation in 'risky behaviour' [32]. Tests of theory would involve manipulation of state-variables to quantify whether an individual's risky behaviour changed in the direction predicted by theory. However, the validity of the empirical test would hinge critically on whether the assayed behaviour did indeed represent a risky behaviour. Researchers may thus inappropriately interpret empirical tests of a given theoretical model because they did not measure the target character [33,34]. The usefulness of the proposed framework is further illustrated by our empirical example: if we had only measured the amount of songs produced as a proxy of aggressiveness, we could have arrived at the conclusion that the more songs produced the more aggressive was the response. Our empirical example implied that more aggressive displays were, in contrast, characterized by a lower—not higher—number of produced songs (figure 4).

Other fields of biology have, notably, been pioneers in some elements of our proposed approach. Specifically, human personality psychology has a long history of focusing on latent variables in the study of behaviour [14], where techniques such as the 'multi-trait multi-method' approaches [35] are commonly used to examine the validity of measurements of latent variables. Nevertheless, key characteristics of behavioural characters, such as within-individual variation owing to adaptive responses to the environment (i.e. 'individual plasticity') and its multi-variate extension (i.e. 'integration of plasticity' [18]), are not fully embedded in human personality research. The treatment of within-individual variation as personality 'signatures' [36,37] in psychology does not, in our reading, appear to be based on evolutionary principles. By contrast, within-individual variation owing to adaptive individual plasticity represents a key concept in evolutionary biology [16]. A possible reason for this mismatch could be the prevailing type of experimental design in human psychology [37], where individuals (or their peers) are typically—though not always—subjected to questionnaires that asks about the subject's typical behaviour (i.e. average, long-term response) in a diverse range of (social and non-social) situations. Our proposed approach would instead require *repeated* exposure to the same questionnaire (over an environmental gradient), such that within- and between-individual (co)variances can be estimated explicitly. Fully integrating multi-level (co)variation in characterizing labile characters would, in our view, represent a very fruitful expansion in both evolutionary biology and human psychology research.

(d) The hierarchical structure of behavioural characters

The hierarchical nature of behaviour and other labile phenotypes represents a key aspect of the evolutionary character framework [3]. At the lowest level of the hierarchy are the behaviours observed in a particular context; those may represent expressions of a lower order character. Two of such lower order characters were evident in the great tit dataset (i.e. 'aggressiveness during laying' and 'aggressiveness during incubation'; figure 4). If such lower order characters represented evolutionary modules with partial—though not full—overlap in function, they should in turn be partly underpinned by a higher order (i.e. context general) character. Indeed, the positive correlation between the two context-specific latent variables supported the existence of such a higher order character (figure 4), which we might (objectively) call 'aggressiveness during the reproductive season'. One could readily extend this approach by including agonistic behaviours expressed in other contexts, for example those expressed outside the reproductive season (e.g. winter dominance interactions). This would yield insight in the generality versus (seasonal) specificity of aggressiveness as a character.

Inclusion of observed behaviours expressed in related but functionally distinct contexts would help to reveal the existence of higher order behavioural characters. For example, a higher order behavioural character representing 'willingness to take risk' might modulate lower order characters, such as aggressiveness, anti-predator boldness and exploratory tendency. An analogy in human psychology—a field that fully acknowledges hierarchical structuring—would be that 'orderliness', 'achievement striving' and 'cautiousness' are all part of a broad factor known as the personality axis 'conscientiousness' [38]. We illustrated this idea empirically by testing whether or not aggressiveness and activity in a novel environment were associated (see electronic supplementary material, table S4). This was not the case, implying that aggressiveness during the reproductive season was quasi-independent of activity in a novel environment, suggesting that the postulated higher order character did not exist in this case. Nevertheless, even if distinct modules (characters) would underpin behaviour in different functional contexts, correlations among them may be observed (cf. 'behavioural syndromes'; [39]). Functionally unrelated behavioural characters might also share proximate mechanism owing to the redundancy in expression pathways [6] resulting in an overarching modularity driven by constraints in the architecture of behaviour rather than functional coherence.

(e) The adaptive nature of behavioural characters

As detailed in this paper, the functional coherence that defines a 'behavioural character' comes with predictions about the (multi-level) structure of behavioural (co)variation. Implicit to the framework is also the adaptive nature of modules, an assumption that can be tested empirically. Specifically, if the organism indeed benefits from functional units in the execution of a particular task (e.g. grabbing objects in our example of the human hand), we explicitly expect natural selection to favour correlations ('correlational selection'; [40]) between the expressed observables (i.e. length of the five fingers). In the case of aggressiveness, we would thus expect strong correlational selection to act on the agonistic behaviours during egg production when ineffective displays, for example calling but not approaching, might have important fitness costs (for example, risk of paternity loss). While awaiting formal phenotypic selection analyses applied to our data, the structure of behaviour is in line with this notion: our point estimates of behavioural correlations were tighter during egg laying compared with incubation (table 2), and the latent variable 'aggressiveness during laying' explained more variance in the agonistic behaviours than its counterpart during incubation (figure 4).

(f) Estimating behavioural character values

Researchers are continuously faced with the challenge of which behavioural data to incorporate in their analyses. What guidelines might one apply once the behavioural character concept has empirically been confirmed? We see two options. First, one could calculate a composite score derived from the structure of the latent variable. In the electronic supplementary material, appendix S4, we detail how an individual's score for the latent variable might be calculated (see also the electronic supplementary material, figure S2). This represents a more appropriate version of the traditionally recommended usage of composite scores from PCA within behavioural ecology [41], while having the advantage of (i) being able to deal with missing data [13] and (ii) avoiding failure to acknowledge the statistical non-independence of repeated measures data [42]. Unfortunately, the usage of such latent scores for further analyses is without doubt more complex and in some circumstances may demand large sample sizes [17,42]. Researchers might, therefore, alternatively use a single observable that closely predicts the behavioural character under study. Of course, such an approach would represent a less precise way of quantifying the character, but would also, logistically and technically, be less challenging. No matter which approach is chosen, it is important to acknowledge the distinction between behavioural characters and the behavioural observables. In some cases, the observable will accurately reflect the target behavioural character, though observables may represent expressions of multiple characters. Above all, we recommend that behavioural characters are defined explicitly in reference to a specific biological process and that behavioural observables should thus be labelled as objectively as possible. Doing so would help to avoid subjectivity in studying behavioural characters [43].

5. Conclusion

Our proposed framework attempts to unite advances in different fields of research in the study of characters. Our framework integrates cross-disciplinary research paradigms, including the study of latent variables in human psychology, the multi-level approach in the study of labile characters in behavioural and evolutionary ecology and the conceptualization of phenotypic organization in evolutionary biology. Such a holistic framework will enhance our ability to characterize the structure of behaviour, and other labile characters, and place it firmly in the realm of evolutionary biology.

Acknowledgements. We thank Marion Nicolaus, Kimberly Mathot, Alexia Mouchet, Ariane Mutzel, Anne-Lise Olsen, Jan Wijmenga, Erica Stuber, and all field assistants and students for help in data collection and discussion. Members of the Department of Behavioural Ecology and Evolutionary Genetics of the Max Planck Institute for Ornithology, Pim Edelaar, Hannah Fried-Petersen, Barbara Helm, Kimberly Mathot, Lars Penke, Alexander Weiss, Dave Westneat, Jan Wijmenga, Jon Wright and two anonymous reviewers, are gratefully acknowledged for their feedback on the manuscript, and Ned Dochtermann, Shinichi Nakagawa, Mihai Valcu, for statistical advice.

Data accessibility. Data were deposited in the dryad repository (www.datadryad.org): doi:10.5061/dryad.1c700.

Funding statement. The study was supported by the German Academic Exchange Service (DAAD), the Max Planck Society and the International Max Planck Research School for Organismal Biology.

References

1. Gould SJ, Lewontin RC. 1979 The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc. R. Soc. Lond. B* **205**, 581–598. (doi:10.1098/rspb.1979.0086)
2. Pigliucci I, Kaplan I. 2000 The fall and rise of Dr Pangloss: adaptationism and the Spandrels paper 20 years later. *Trends Ecol. Evol.* **15**, 66–70. (doi:10.1016/S0169-5347(99)01762-0)
3. Wagner GP. 2001 *The character concept in evolutionary biology*. San Diego, CA: Academic Press-Elsevier Science.
4. Duckworth RA. 2008 The role of behavior in evolution: a search for mechanism. *Evol. Ecol.* **23**, 513–531. (doi:10.1007/s10682-008-9252-6)
5. Wagner GP, Stadler PF. 2003 Quasi-independence, homology and the unity of type: a topological theory of characters. *J. Theor. Biol.* **220**, 505–527. (doi:10.1006/jtbi.2003.3150)
6. Wagner GP, Pavlicev M, Cheverud JM. 2007 The road to modularity. *Nat. Rev. Genet.* **8**, 921–931. (doi:10.1038/nrg2267)
7. Schwenk K. 2001 Functional units and their evolution. In *The character concept in evolutionary biology* (ed. G Wagner), pp. 156–198. San Diego, CA: Academic Press.
8. Wake MH. 2008 Organisms and organization. *Biol. Theory* **3**, 213–223. (doi:10.1162/biot.2008.3.3.213)
9. Susman RL. 1979 Comparative and functional morphology of hominoid fingers. *Am. J. Phys. Anthropol.* **50**, 215–236. (doi:10.1002/ajpa.1330500211)
10. Krebs JR. 1982 Territorial defence in the great tit: do residents always win? *Behav. Ecol. Sociobiol.* **11**, 185–194. (doi:10.1007/BF00300061)
11. Pigliucci M. 2003 Phenotypic integration: studying the ecology and evolution of complex phenotypes. *Ecol. Lett.* **6**, 265–272. (doi:10.1046/j.1461-0248.2003.00428.x)
12. Houle D. 2001 Characters as the units of evolutionary change. In *The character concept in evolutionary biology* (ed. GP Wagner), pp. 109–140. San Diego, CA: Academic Press.
13. Gelman A, Hill J. 2007 *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University Press.
14. Bollen KA. 2002 Latent variables in psychology and the social sciences. *Annu. Rev. Psychol.* **53**, 605–634. (doi:10.1146/annurev.psych.53.100901.135239)
15. Dingemanse NJ, Kazem AJN, Réale D, Wright J. 2010 Behavioural reaction norms: animal personality meets individual plasticity. *Trends Ecol. Evol.* **25**, 81–89. (doi:10.1016/j.tree.2009.07.013)
16. Nussey DH, Wilson AJ, Brommer JE. 2007 The evolutionary ecology of individual phenotypic plasticity in wild populations. *J. Evol. Biol.* **20**, 831–844. (doi:10.1111/j.1420-9101.2007.01300.x)
17. Dingemanse NJ, Dochtermann NA. 2013 Quantifying individual variation in behaviour: mixed-effect modelling approaches. *J. Anim. Ecol.* **82**, 39–54. (doi:10.1111/1365-2656.12013)
18. Pigliucci M. 2001 Characters and environments. In *The character concept in evolutionary biology* (ed. G Wagner), pp. 365–390. San Diego, CA: Academic Press.
19. Birkhead T, Atkin L, Møller A. 1987 Copulation behaviour of birds. *Behaviour* **101**, 101–138. (doi:10.1163/156853987X00396)
20. Roff DA. 1997 *Evolutionary quantitative genetics*. New York, NY: Chapman & Hall.
21. Nicolaus M, Tinbergen JM, Bouwman KM, Michler SPM, Ubels R, Both C, Kempenaers B, Dingemanse NJ. 2012 Experimental evidence for adaptive personalities in a wild passerine bird. *Proc. R. Soc. B* **279**, 4885–4892. (doi:10.1098/rspb.2012.1936)
22. Hulbert SH. 1984 Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr.* **54**, 187–211. (doi:10.2307/1942661)
23. Nakagawa S, Schielzeth H. 2010 Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biol. Rev. Camb. Philos. Soc.* **85**, 935–956. (doi:10.1111/j.1469-185X.2010.00141.x)
24. Inman H, Bradley E. 1989 The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. *Commun. Stat. Theory Methods* **18**, 3851–3874. (doi:10.1080/03610928908830127)
25. Stuber EF, Araya-Ajoy YG, Mathot KJ, Mutzel A, Nicolaus M, Wijmenga JJ, Mueller JC, Dingemanse NJ. 2013 Slow explorers take less risk: a problem of sampling bias in ecological studies. *Behav. Ecol.* **24**, 1092–1098. (doi:10.1093/beheco/art035)
26. R Core Team. 2012 R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
27. Hadfield JD. 2010 MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R Package. *J. Stat. Softw.* **33**, 1–36.
28. Ellison AM. 2004 Bayesian inference in ecology. *Ecol. Lett.* **7**, 509–520. (doi:10.1111/j.1461-0248.2004.00603.x)
29. Fox J. 2006 Structural equation modeling with the sem package in R. *Struct. Equ. Model.* **13**, 465–486. (doi:10.1207/s15328007sem1303_7)
30. Armbruster W, Schwaegerle K. 1996 Causes of covariation of phenotypic traits among populations. *J. Evol. Biol.* **9**, 261–276. (doi:10.1046/j.1420-9101.1996.9030261.x)
31. Owens IPF. 2006 Where is behavioural ecology going? *Trends Ecol. Evol.* **21**, 356–361. (doi:10.1016/j.tree.2006.03.014)
32. Wolf M, van Doorn GS, Leimar O, Weissing FJ. 2007 Life-history trade-offs favour the evolution of animal personalities. *Nature* **447**, 581–584. (doi:10.1038/nature05835)
33. Watanabe NM, Stahlman WD, Blaisdell AP, Garlick D, Fast CD, Blumstein DT. 2012 Quantifying personality in the terrestrial hermit crab: different measures, different inferences. *Behav. Process.* **91**, 133–140. (doi:10.1016/j.beproc.2012.06.007)
34. Carter AJ, Marshall HH, Heinsohn R, Cowlishaw G. 2012 How not to measure boldness: novel object and antipredator responses are not the same in wild baboons. *Anim. Behav.* **84**, 603–609. (doi:10.1016/j.anbehav.2012.06.015)
35. Campbell DT, Fiske DW. 1959 Convergent and discriminant validation by the multitrait-mulit method matrix. *Psychol. Bull.* **56**, 81–105. (doi:10.1037/h0046016)
36. Mischel W, Shoda Y. 1995 A cognitive-affective system theory of personality: reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychol. Rev.* **102**, 246–268. (doi:10.1037/0033-295X.102.2.246)
37. Penke L, Denissen J, Miller G. 2007 The evolutionary genetics of personality. *Eur. J. Pers.* **587**, 549–587. (doi:10.1002/per.629)
38. Jang KL, McCrae RR, Angleitner A, Riemann R, Livesley WJ. 1998 Heritability of facet-level traits in a cross-cultural twin sample: support for a hierarchical model of personality. *J. Pers. Soc. Psychol.* **74**, 1556–1565. (doi:10.1037/0022-3514.74.6.1556)
39. Sih A, Bell A, Johnson JC. 2004 Behavioral syndromes: an ecological and evolutionary overview. *Trends Ecol. Evol.* **19**, 372–378. (doi:10.1016/j.tree.2004.04.009)
40. Sinervo B, Svensson E. 2002 Correlational selection and the evolution of genomic architecture. *Heredity* **89**, 329–338. (doi:10.1038/sj.hdy.6800148)
41. McGregor PK. 1992 Quantifying responses to playback: one, many, or composite multivariate measures? In *Playback studies of animal communication* (ed. PK McGregor), pp. 79–96. New York, NY: Plenum Press.
42. Budaev SV. 2010 Using principal components and factor analysis in animal behaviour research: caveats and guidelines. *Ethology* **116**, 472–480. (doi:10.1111/j.1439-0310.2010.01758.x)
43. Carter AJ, Feeney WE, Marshall HH, Cowlishaw G, Heinsohn R. 2012 Animal personality: what are behavioural ecologists measuring? *Biol. Rev. Camb. Philos. Soc.* **88**, 465–475. (doi:10.1111/brv.12007)

Chapter 2

Multi-level analysis of reaction norms: an approach to estimate short-term, long-term and reaction norm repeatability

Yimen G. Araya-Ajoy, Kim J. Mathot, Niels J. Dingemanse

Submitted to Methods in Ecology and Evolution (2nd revision)

Abstract

1. Evolutionary ecologists increasingly study reaction norms that are expressed repeatedly within the same individual's lifetime. For example, foragers continuously alter anti-predator vigilance in response to moment-to-moment changes in predation risk. Variation in this form of plasticity occurs both among and within individuals. Among-individual variation in plasticity (cf. individual by environment interaction or I×E) is commonly studied; by contrast, despite increasing interest in its evolution and ecology, within-individual variation in phenotypic plasticity is not.
2. We propose a study design and statistical approach (based on repeated measures and multi-level random regression modelling) that enables the study of variation in reaction norms at different hierarchical levels (such as among- and within-individuals). The approach enables the calculation of repeatability of reaction norm intercepts (cf. average phenotype) and slopes (cf. level of phenotypic plasticity); these indices are not specific to measurement or scaling and are readily comparable across data sets.
3. The proposed framework also enables calculation of repeatability at different temporal scales (such as short- and long-term repeatability) thereby answering calls for the development of approaches enabling scale-dependent repeatability calculations.
4. We introduce a simulation package in the R statistical language to assess power, imprecision and bias for multi-level random regression that may be utilised for realistic datasets (cf. unequal sample sizes across individuals, missing data, etc).
5. We apply the method to a worked example to illustrate its utility. We conclude that consideration of multi-level variation in reaction norms deepens our understanding of the hierarchical structuring of labile characters and helps reveal the biology in heterogeneous patterns of within-individual variance that would otherwise remain ‘unexplained’ residual variance.

Introduction

Patterns of individual variation in labile phenotypic characters, such as behaviour, physiology and life-history, are increasingly studied as 'reaction norms' (Nussey et al. 2007; Dingemanse et al. 2010b; Westneat et al. 2011), functions relating individual phenotypes to environmental variables (Schlichting & Pigliucci 1998). Individual reaction norms measure two distinct components of the phenotype, where each individual is characterised by a certain combination of reaction norm intercept and slope (Fig. 1). The former might, for example, represent the individual's average phenotype expressed in a mean-centered environment, the latter its level of phenotypic plasticity (Nussey et al. 2007). Evolutionary ecologists routinely investigate whether individuals differ in reaction norm slope (i.e. 'individual by environment interaction'; I×E) because it provides information about the potential for heritable (i.e. evolvable) variation in phenotypic plasticity (Nussey et al. 2007). Reaction norm approaches are also increasingly applied in other fields, such as behavioural ecology (Dingemanse et al. 2010b; Westneat et al. 2014a), and endocrinology (Dingemanse et al. 2010a; Lema & Kitano 2013).

Populations often consist of animals that differ in level of phenotypic plasticity (Nussey et al. 2007; Mathot et al. 2011; Forsman 2014). For example, individuals are distinct in how their reproductive profiles change with age (Nussey et al. 2006) or how strongly they adjust the timing of their reproduction to spring temperature (Brommer et al. 2008). Until recently, many ecological studies focused on phenotypic characters that are expressed a small number of times in the life-time of the individual, such as lay date or clutch size in short-lived passerine birds (van de Pol 2012). In the past few years, however, evolutionary ecologists increasingly concentrate on phenotypic characters that are expressed many times in the life-time of the individual, and are adjusted to environmental variables that vary within individuals over relatively short temporal scales. For example, parents adjust the inter-visit interval between subsequent visits to their nest as a function of information about nestling state (e.g. begging intensity) acquired during the previous visit (Wright et al. 2010), and animals often alter their anti-predator vigilance in response to moment-to-moment changes in perceived predation risk by avian predators (Mathot et al. 2011). Individual differences in such short-term adjustments of behaviour constitute a major of topic of interest in behavioural neurophysiological research (Koolhaas et al. 2010), where it has been proposed that 'responsiveness' (Wolf et al. 2008) shows individual repeatability both across time and

functional contexts due to fundamental differences in how individuals process environmental input (Coppens et al. 2010; Mathot et al. 2012).

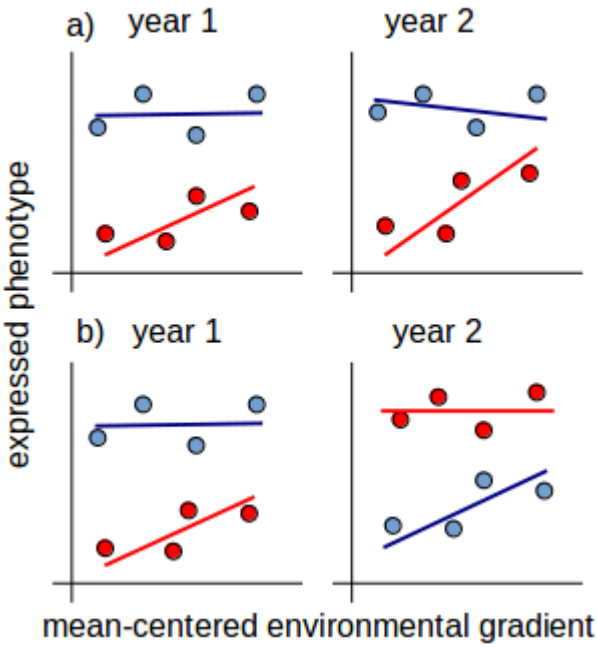

Figure 1. *Relationships between phenotypes (dots) and environmental conditions. In each graph, we show a reaction norm for each of two individuals each assayed once in each of five environmental conditions within the same year; lines represent each individual's reaction norm, blue symbols are for individual A and red symbols for individual B. The environmental gradient is scaled (range -0.5 to 0.5, i.e standardised to two standard deviation units) following Gelman (2008) and could either represent a continuous environmental gradient or a two-level factor. In all graphs, individuals differ in reaction norm slopes and intercepts within each year and both reaction norms components also change within individuals from one year to the next. At the same time, both reaction norm components either (a) do or (b) do not show cross-year individual repeatability.*

Individual variation in responsiveness can be adaptive both at the among-individual and within-individual level (Dingemanse & Wolf 2013). That is, the specific environmental conditions faced by an individual at a particular point in time may constrain, limit, or affect the balance between costs and benefits of phenotypic plasticity (Auld et al. 2010). Within-individual variation in reaction norms can exist in nature because reaction norms are often 'multidimensional', i.e. an individual's phenotype may respond to multiple environmental axes (Westneat et al. 2009). In cases where such reaction norm 'planes' are warped because of interacting effects of environmental axes (sensu Westneat et al. 2011), the level of phenotypic plasticity with regard to one environmental axis varies as a function of another. In bird species with bi-parental care, for example, provisioning rate increases within individual

parents as a function of nestling age but the slope of this repeatedly expressed reaction norm varies as a function of the nest's brood size (Westneat et al. 2011). Repeatedly expressed reaction norms have also become a focus in behavioural ecology, a field that is currently developing adaptive theory for the evolution of repeatable—vs. unrepeatable—variation in phenotypic plasticity (Wolf et al. 2008; McNamara et al. 2009). Empirical research will thus increasingly focus on the estimation of multi-level variation in reaction norms, such as plastic responses that vary among vs. within individuals or populations (Briffa et al. 2008; Westneat et al. 2011; Dingemanse & Wolf 2013).

In this paper, we propose a study design and statistical approach that researchers may apply to estimate variation in reaction norm parameters across multiple hierarchical levels. Our proposed multi-level approach is suitable for cases where individuals are exposed to the same environmental gradient multiple times in their lives (examples given above), opening the possibility to study among vs. within-individual variation in reaction norm parameters. The ability to differentiate between repeatable vs. unrepeatable variation in phenotypic plasticity constitutes an important means of testing predictions of current adaptive theory (Dingemanse & Wolf 2013; Westneat et al. 2014b). Theoreticians, for example, predict that competition for resources in a heterogeneous environment promotes the emergence of an evolutionary stable mix of plastic vs. non-plastic sampling strategies (Wolf et al. 2008). Importantly, individual repeatability in plasticity is only expected in cases where the endogenous and exogenous features of organisms affecting the costs and benefits of plasticity are stable over time. If this condition is not met, selection could just as likely favor individuals playing mixed strategies at the evolutionary stable strategy equilibrium (Wolf et al. 2008). Therefore, repeatable vs. unrepeatable variation in reaction norms is expected to vary as a function of species-, population- or environment-specific ecological conditions.

Multi-level analysis of variation in reaction norms

We detail how variation in reaction norm components can be partitioned across multiple levels in two steps. First, we discuss the so-called 'phenotypic equation' that evolutionary ecologists routinely use to estimate individual variation in reaction norms (Nussey et al. 2007; Dingemanse et al. 2010b; Westneat et al. 2011). Second, we detail how multi-level implementations of this equation enable the estimation of variation within and among individuals in reaction norm components, from which estimates of repeatability may

subsequently be calculated using established approaches (Nakagawa & Schielzeth 2010)

Individual variation in reaction norms is—in its simplest form—modelled using a random regression mixed-effect model (Nussey et al. 2007), which we present in the following phenotypic equation (Eqn. 1a):

$$y_{ik} = (\beta_0 + ind_{0k}) + (\beta_1 + ind_{1k}) x_{ik} + e_{0ik} \quad \text{(Eqn. 1a)}$$

Here, a single phenotypic response (y_{ik}), such as the level of aggressiveness by individual *k* exhibited at instance *i* is modelled as a function of x_{ik} , a covariate or factor representing for example the breeding stage of individual *k* at instance *i*. This phenotypic response (y_{ik}) may be described by five distinct elements: i) the population-mean reaction norm intercept (β_0 ; the grand mean value of average individual responses), ii) the population-mean reaction norm slope (β_1 ; the coefficient relating x_{ik} to y_{ik}), iii) the individual's deviation in reaction norm intercept (ind_{0k}) from the population-mean intercept (β_0), iv) the individual's deviation in reaction norm slope (ind_{1k}) from the population-mean slope (β_1), and v) the instance's deviation from the individual's reaction norm (e_{0ij}). This model is called a 'random regression' because the individual-specific deviations from the population-mean value with respect to intercepts (ind_{0k}) and slopes (ind_{1k}) are typically assumed to be 'drawn' from (i.e. follow) a bivariate (normally Gaussian; MVN) distribution with a mean of zero and covariance matrix (Ω_{ind}) to be estimated from the data. The (co)variances for this distribution are defined by the variance in intercepts among individuals (V_{ind_0}), the variance in slopes among individuals (V_{ind_1}), and the covariance between intercepts and slopes (Cov_{ind_0, ind_1} ; also commonly expressed as a correlation: $r_{ind_0, ind_1} = Cov_{ind_0, ind_1} / \sqrt{V_{ind_0} V_{ind_1}}$). The deviations from individual reaction norms for each instance (e_{0ij}) are also modelled (again, typically assuming a Gaussian distribution) with a mean of zero and an estimated residual variance (V_{e_0}) (Eqn. 1b):

$$\begin{bmatrix} ind_{0_k} \\ ind_{1_k} \end{bmatrix} MVN\left(0, \Omega_{ind}\right) \quad : \quad \Omega_{ind} = \begin{bmatrix} V_{ind_0} & Cov_{ind_0, ind_1} \\ Cov_{ind_1, ind_0} & V_{ind_1} \end{bmatrix} \qquad \text{(Eqn. 1b)}$$

$$\left[e_{0ik}\right] N\left(0, \Omega_e\right) \quad : \quad \Omega_e = \left[V_{e_0}\right]$$

Unfortunately, estimates of variance in reaction norm components are influenced by how the focal environmental gradient (x_{ijk}) is measured and scaled (Schaeffer 2004; Gelman 2008), both in terms of magnitude and sign. This is in part because V_{ind_0} represents the variance among individuals in intercept value, which is conditional to the positioning of the intercept along the environmental axis (x_{ik}) in cases where individuals differ in phenotypic plasticity (i.e. $V_{ind_1} > 0$). Hence, the choice on whether and how to center the environmental axis (x_{ik}) represents an important decision (Enders & Tofighi 2007). In the study of reaction norms, environmental covariates (x_{ik}) are typically centered on their mean value, such that V_{ind_0} represents the among-individual variance in intercepts in the average environmental condition; this decision makes intuitive sense in part because it facilitates cross-study comparisons (Nussey et al. 2007; Dingemanse & Dochtermann 2013). Throughout, we assume in our verbal descriptions of variance components that fixed effects were mean-centered and standardised to two standard deviation units (for a full discussion, see (Gelman 2008)).This transformation may also be applied to two-level factors (e.g. low vs. high, before vs. after, control vs. treated), a commonly used setup in ecological and evolutionary studies, and the transformed variable fitted as a covariate (x_{ik} into the model, as we do in the worked example and simulated scenario detailed below).

The random regression detailed in Eqn.1 explicitly estimates a single reaction norm intercept and slope for each individual. However, here we are concerned with a biological scenario where an individual expresses the same reaction norm multiple times (Fig. 1), and it is this repeated measures structure in the data that provides the opportunity to study variation in reaction norms among *and* within individuals. For example, red knots (*Calidris canutus*) decrease the size of their gizzard during migration (Piersma & Drent 2003) and individuals are therefore repeatedly exposed to this seasonal variation in every *year* that they survive and migrate. Great tits (*Parus major*) adjust their aggressive response to changes in the breeding stage of their female (Araya-Ajoy & Dingemanse 2014), and individuals experience this type of variation in their social environmental every year. We propose here that we can use these

repeated expressions of an individual's reaction norm to partition variation in reaction norm intercepts and slopes across hierarchical levels. We may do this by creating an extra random effect that groups the phenotypic expressions in response to an environmental gradient for a particular individual in a unit capturing the temporal dependency of observations. We call this extra random effect 'series' to denote a period of time (e.g. year or day) within which one managed to obtain phenotypic data over a range of environmental conditions for the same individual. For example, in the case where each individual great tit male has to adjust his aggressiveness to the breeding stage of its mate every year, 'series' would represent a unique combination of breeding year and individual. Similarly, series would represent a unique combination of year and individual in a study of gizzard size in red knots. These examples illustrate that for any individual with repeated series one can obtain information about series-specific reaction norm components (Fig. 1). The overall variance in reaction norm components among all collected series in the dataset consequently represents the phenotypic variance in reaction norms to be partitioned across hierarchical levels. Such partitioning is logically only warranted in cases where the among-series variance in reaction norms is nonzero. This pre-condition may be tested by substituting the random effect 'individual' for 'series' in equation 1. Provided that among-series variance was indeed present in this model, the classic phenotypic equation (Eqn. 1) may then be expanded by including random intercepts and slopes for *both* individual and series identity (Eqn. 2a):

$$y_{ijk}=\left(\beta_0+ind_{0k}+series_{0jk}\right)+\left(\beta_1+ind_{1k}+series_{1jk}\right)x_{ijk}+e_{0ijk} \quad \text{(Eqn. 2a)}$$

Here, we now partition a single phenotypic response (y_{ijk}) by individual *k* exhibited at series *j* at instance *i* as a function of x_{ijk} . Adding random intercepts and slopes for series identity thereby enables us to estimate both the individual-mean reaction norm intercept $\left(\beta_0+ind_{0k}\right)$ and slope $\left(\beta_1+ind_{1k}\right)$ over all its series (e.g. days or years), as well as each series' deviation from the individual's mean reaction norm intercept ($series_{0jk}$) and slope ($series_{1jk}$). This model will enable us to directly estimate the variance among individuals in average reaction norm intercept (V_{ind_0}) and slope (V_{ind_1}) as well as the within-individual among-series variance in those intercepts (V_{series_0}) and slopes (V_{series_1}) (Eqn 2b):

$$\begin{bmatrix} ind_{0_k} \\ ind_{1_k} \end{bmatrix} MVN\left(0, \Omega_{ind}\right) \quad : \quad \Omega_{ind} = \begin{bmatrix} V_{ind_0} & Cov_{ind_0, ind_1} \\ Cov_{ind_1, ind_0} & V_{ind_1} \end{bmatrix} \qquad \text{(Eqn 2b)}$$

$$\begin{bmatrix} series_{0_{jk}} \\ series_{1_{jk}} \end{bmatrix} MVN\left(0, \Omega_{series}\right) \quad : \quad \Omega_{ind} = \begin{bmatrix} V_{series_0} & Cov_{series_0, series_1} \\ Cov_{series_1, series_0} & V_{series_1} \end{bmatrix}$$

$$\left[e_{0ijk}\right] N\left(0, \Omega_e\right) \qquad \Omega_e = \left[V_{e_0}\right]$$

The presence of repeated measures for intercepts and slopes of the same individuals consequently enables the calculation of a standardised index (i.e. repeatability) that represents the proportion of variance in a focal reaction norm component among all series that is explained by differences among individuals. Following established approaches for mixed-effect models, this individual repeatability may be calculated for reaction norm intercepts (Eqn. 3a), slopes (Eqn. 3b) and the reaction norms as a whole (Eqn. 3c):

$$R_{intercept} = \frac{V_{ind_0}}{V_{ind_0} + V_{series_0}} \qquad \text{(Eqn. 3a)}$$

$$R_{slope} = \frac{V_{ind_1}}{V_{ind_1} + V_{series_1}} \qquad \text{(Eqn. 3b)}$$

These standardised proportions, notably, represent the repeatability of an individual's estimated reaction norm intercept (i.e. repeatability of average behaviour), and reaction norm slope, and thus exclude residual error (V_{e_0}). A graphical illustration of the general idea is given in Fig. 1, where the reaction norms of two hypothetical individuals were measured in each of two years. Two scenarios are given. In the first scenario, there is variation in reaction norms among all series (cf. among all unique combinations of individual and year) with individuals showing non-zero cross-year repeatability in intercepts and slopes (Fig. 1a). In the second scenario, there is also variation in reaction norms among all series but individuals do not show cross-year repeatability in intercept or slope (Fig 1b). As mentioned above, it is worth noting that the magnitude and sign of reaction norm components are statistically influenced by the measurement and scaling of the focal environmental gradient, hampering comparability across data sets (Schaeffer 2004; Gelman 2008). This is, importantly, not the case for our estimates of reaction norm repeatability. This is because estimates of equations

3a and 3b are proportions of a single variance component which are scaled in the same way across all levels, thereby cancelling out any biasing effects of centering or scaling, making them comparable across datasets for future meta-analyses.

Intercept repeatability derived from equation 3a refers to the repeatability of an individual's average phenotype expressed over all its series (i.e. in the average environment if mean-centered environmental gradients were modelled). This is distinctly different from what is generally referred to as repeatability (e.g. individual repeatability), which would additionally include the residual variance (V_{e_0}) in the denominator of equation 3a, resulting in the commonly used formula to estimate repeatability of a repeatedly expressed phenotype (detailed in our General discussion below). Note that it is possible to also include the residual variance in equation 3b, but the interpretation of this ratio needs to be taken with care because residual variance stands for the extent to which observed phenotypes deviate from that predicted by the statistical model, whereas slope variance represents variation in how individuals differ in their response to an environmental gradient, and is thus measured in a different unit.

Analogous to what is termed adjusted repeatability (Nakagawa & Schielzeth 2010), one can calculate an adjusted slope repeatability. Researchers may be interested in accounting for population-level effects of environmental variables that induce variation in a nested level (Schielzeth & Nakagawa 2013), here, within-individual variation in plasticity. Specifically, within-individual among-series variance in reaction norm slopes would occur when the average individual in the dataset modifies its level of phenotypic plasticity as a function of environmental factors that vary at the series level. In our worked example (detailed below) this occurs because series are defined by the unique combination of individual and age class, and the response of the average individual (to predation risk) varies as a function of age (Fig. 2). Such patterns are inherently caused by multi-dimensional plasticity (Westneat et al. 2009), and happen when the plastic response to an environmental gradient (predation risk) depends on another environmental gradient (age). One can calculate an adjusted slope repeatability that controls for population-average multi-dimensional plasticity by fitting an interaction between the environment gradient that varies within the series (predation risk) and the series-level environmental effect (age class). This could be done either by including an interaction term between two fixed effects (e.g. predation risk × age) or an interaction term between a fixed (predation risk) and random (age category) effect (this decision would

depend on whether the series-level environmental variable was continuous vs. categorical, and, if the latter, whether it harbored few vs. many levels.). When these interaction effects are modelled, within-individual among-series variation in reaction norm slopes will exclusively represent multi-dimensional plasticity due to unknown environmental variables, or unmeasured among-individual variation in multidimensional plasticity. Modelling this population-level interaction term, will also automatically account for biased sampling of individuals across the environmental gradient that varies within the series.

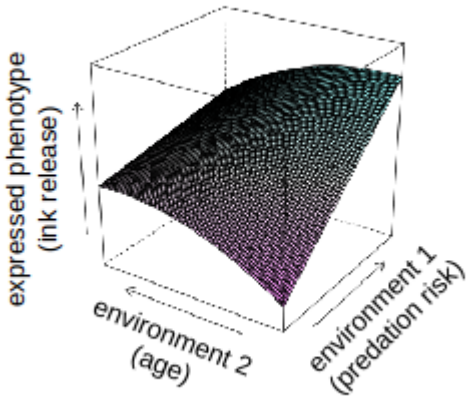

Figure 2. *A two-dimensional reaction norm plot where the expressed phenotype (ink release) varies as a function of the interaction between a within-series (predation risk) and within-individual among-series (age) factor within the average individual (e.g. the plastic response to predation risk depends on age). Here, ink release increases with predation risk but the level of this response decreases with age. Consequently, the level of within-individual plasticity as a function of predation risk would appear to vary across series of the same individual if age was not modelled in the statistical analysis. Within-individual among-series variance in phenotypic plasticity with respect to a single environmental gradient thus occurs when interactive effects of environmental factors are not considered.*

Sampling design

The statistical approach that we advocate requires researchers to measure the response to an environmental gradient repeatedly for the same set of individual subjects. Because our interest is in estimating among-individual variance in reaction norms, within-individual variance in reaction norms, and residual variance, a particular sampling design is strictly required. First, individuals should be assayed at least three times within each series, either once in each of three environmental conditions along a gradient, or twice in each of two (discrete) environmental conditions. This design enables the estimation of residual within-series variance. Second, two or more series are required per individual. This condition

enables the separation of variation in reaction norms into among- and within-individual components. Ideally, all series should have a minimum of three measurements over the environmental gradient (see above); however, series with fewer measurements should not be discarded because such data contribute information to the population and individual level parameters (see e.g. Martin et al. 2011). We use simulations to explore the consequences of different decisions regarding sampling designs (detailed below).

Bias, imprecision and power

Two recent papers detail the optimal sampling designs for parameter estimation in single-level random regression models (Eqn. 1) (Martin et al. 2011; van de Pol 2012). Here, we used data simulations to determine the optimal sampling designs necessary to expand such analyses to include the proposed multi-level scenario (Eqn. 2). We simulated a scenario where the environmental gradient consisted of two levels (e.g. low vs. high predation risk), a common scenario in evolutionary ecology. Simulations were set up such that each individual was assayed twice in each of the contexts within each series; we independently varied the total number of individuals sampled (2-60) and the number of series per individual (2-60). We generated 100 data sets for each combination of *j* individuals assayed in a number of *k* series, for a total of *n* observations. Values were simulated assuming a population-level intercept (β_0) of zero, a population level slope (β_1) of 0.5, and a residual variance (V_{e_0}) of 0.4. Deviations of individual-level intercepts and slopes from population-mean values were simulated assuming a bivariate normal distribution (MVN) with a mean of zero and among-individual covariance matrix (Ω_{ind}) with an intercept variance (V_{ind_0}) of 0.3, slope variance (V_{ind_1}) of 0.1 and an intercept-slope covariance (Cov_{ind_0, ind_1}) of 0.1 (i.e. corresponding to an intercept-slope correlation (r_{ind_0, ind_1}) of 0.5). Deviations of each series from an individual's mean reaction norm intercept and slope were also drawn from a bivariate distribution with a mean of zero and among-series covariance matrix (Ω_{series}) with intercept variance (V_{series_0}) of 0.3, slope variance (V_{series_1}) of 0.1 and covariance ($Cov_{series_0, series_1}$) of 0.1 (i.e. corresponding to an intercept-slope correlation ($r_{series_0, series_1}$) of 0.5). Parameters used to simulate the data were chosen to reflect reasonable values based on published work, but note that multi-level random regression estimates are not common.

We assessed three aspects of the performance of the random regression models; bias, imprecision and power. We defined bias as a quantitative term describing the disagreement between model estimates and the 'true' value. We operationalized this measurement as the absolute difference between the median model estimate for 100 simulated data sets and the value used to generate the data sets. We measured imprecision as the relative coefficient of variation (CV × 100) of the model estimates derived from the 100 datasets sampled for each design (note that lower values indicate greater precision). We measured the power to detect significant among-individual variance in intercepts and slopes, as the proportion of mixed-effect models applied to the 100 simulated data sets that correctly rejected the null hypothesis of zero variance at each of these levels. To assess statistical significance (i.e. p-value), we compared differences in two times the estimated log-likelihood to a chi-square-distribution assuming degrees of freedom equal to the number of constrained parameters. Alternative methods exist for statistical inference; however, we focused on this likelihood ratio test because is widely used (Schaeffer 2004).

Results

Our simulations revealed that the optimal sampling design (for the scenario detailed in section Simulation Procedure) depends on whether researchers are aiming to minimise bias, maximise precision, or maximise power. As a general pattern, parameters at the series level can be estimated with less bias and imprecision for a given sampling design than parameters at the among-individual level (Figs. 3 & 4). For most parameters, a total sample size of 400 with more than 10 sampled individuals enables estimates with low levels of bias (~ 10%). However, reliable estimates for intercept-slope correlations at the among-individual level require larger sample sizes (circa 1000 observations for more than 20 individuals sampled). This particular sample size will also provide a high power (> 0.9) for detecting significant among-individual variation in intercepts and slopes (Fig. 5). By contrast, optimal sampling designs for achieving high precision require markedly larger sample sizes. For example, with 2000 observations and more than 50 sampled individuals, the imprecision in the parameters will range between ~0-30%, where the most imprecisely estimated parameter is the among-individual intercept-slope covariance. An imprecision of 30% in this parameter will mean that the point estimate of a covariance of 0.1 lies between 0.04-0.16 95% of the time.

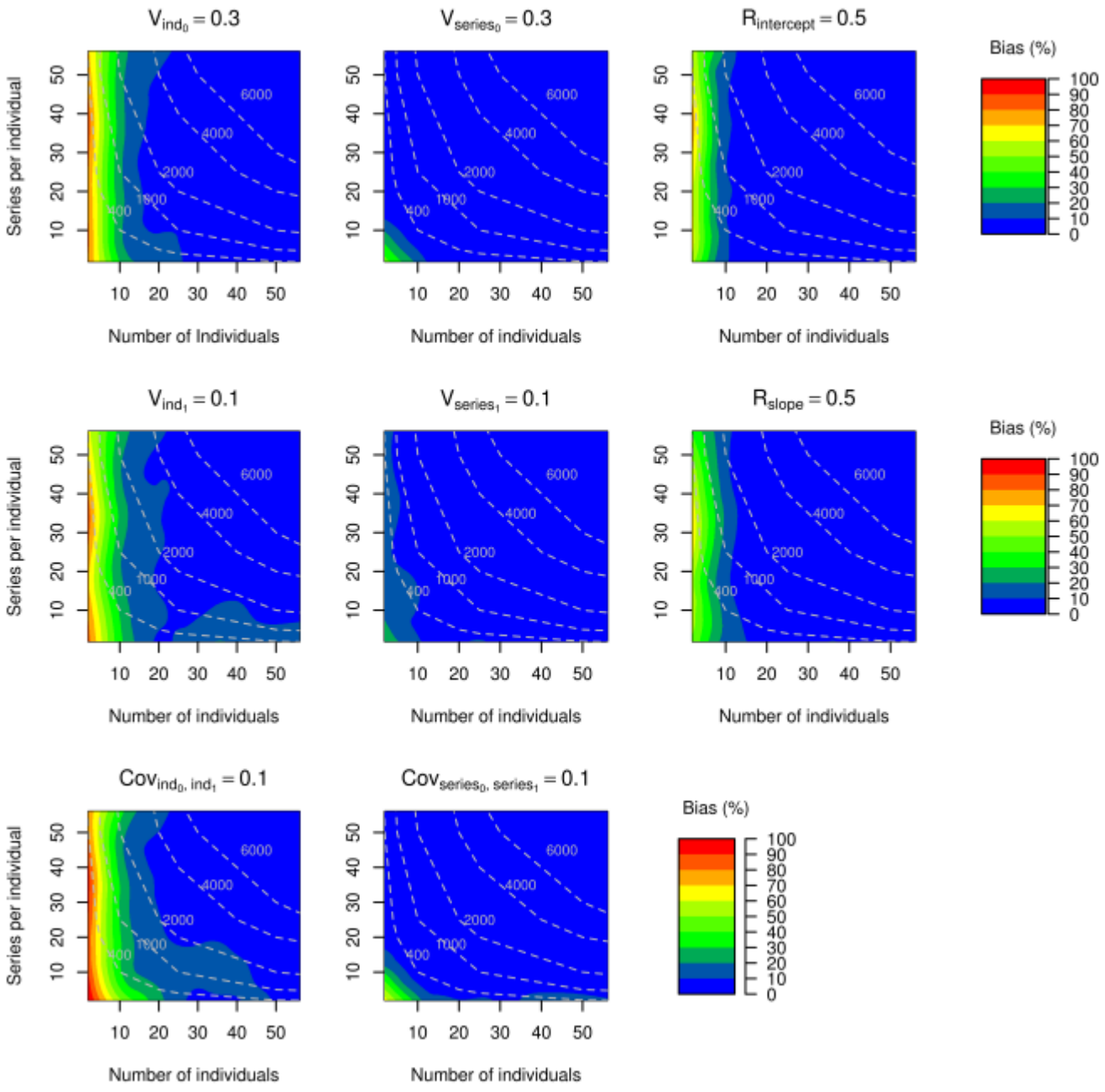

Figure 3. *Relative bias of random regression estimates as a function of the number of individuals sampled and the number of series per individual. Models were applied to simulated data sets with among-individual variance in intercepts (V_{ind_0} =0.1), in slopes (V_{ind_1} =0.05), intercept- slope covariance (cov_{ind_0, ind_1} =0.1) and among-series variance in intercepts (V_{series_0} =0.4), in slopes (V_{series_1} =0.1) and intercept-slope covariance ($cov_{series_0, series_1}$ =0.1). Different colours depict areas between isoclines of similar levels of inaccuracy; isoclines were determined by bilinear interpolation between the sampled integer values of the number of individuals and the number of series per individual.*

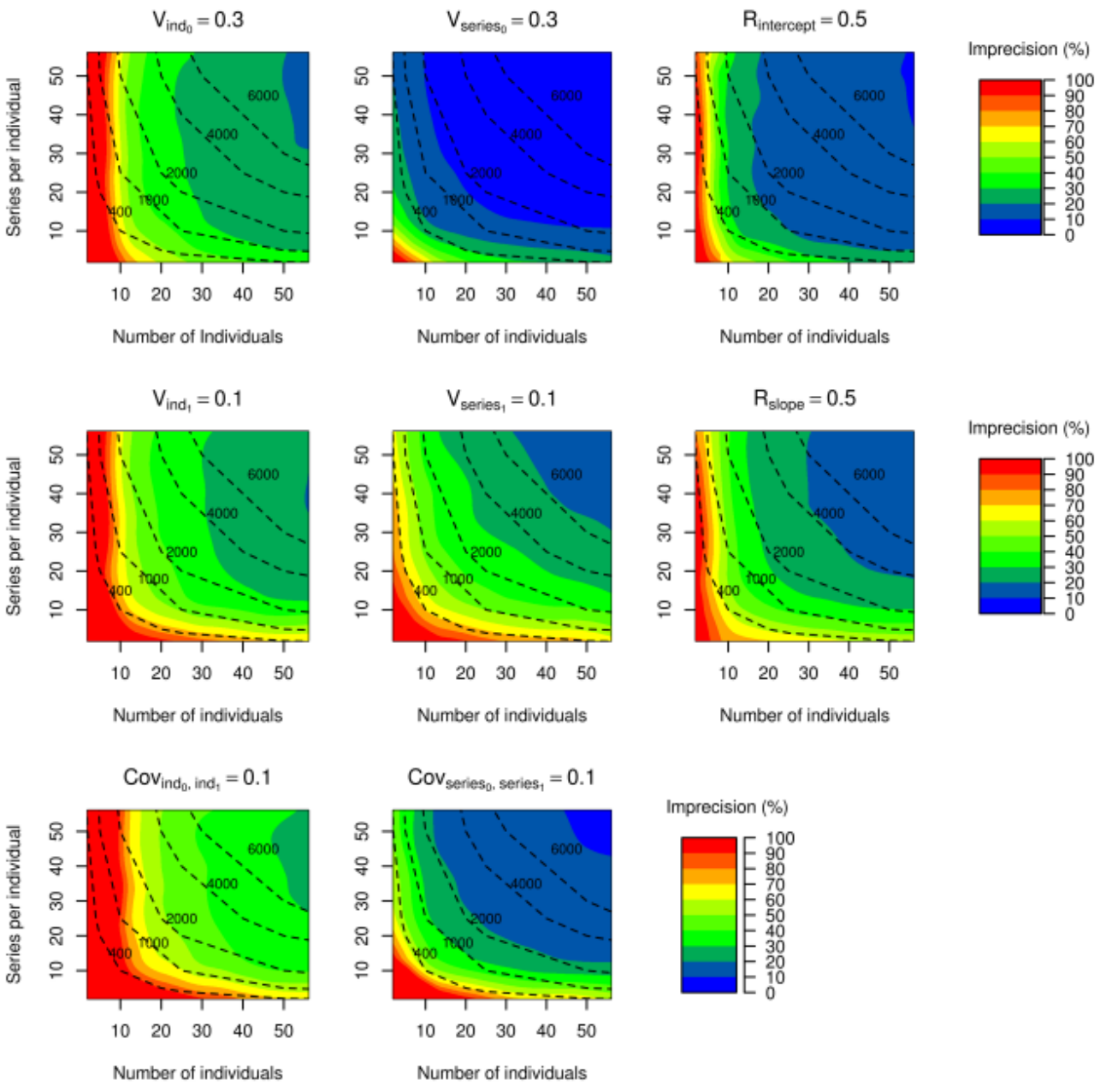

Figure 4. *Imprecision of random regression estimates as a function of the number of individuals sampled and the number of series per individual. Models were applied to simulated data sets with among-individual variance in intercepts (V_{ind_0} =0.1), in slopes (V_{ind_1} =0.05), intercept- slope covariance (cov_{ind_0,ind_1} =0.1) and among-series variance in intercepts (V_{series_0} =0.4), in slopes (V_{series_1} =0.1) and intercept-slope covariance ($cov_{series_0,series_1}$ =0.1). Different colours depict areas between isoclines of similar levels of inaccuracy; isoclines were determined by bilinear interpolation between the sampled integer values of the number of individuals and the number of series per individual.*

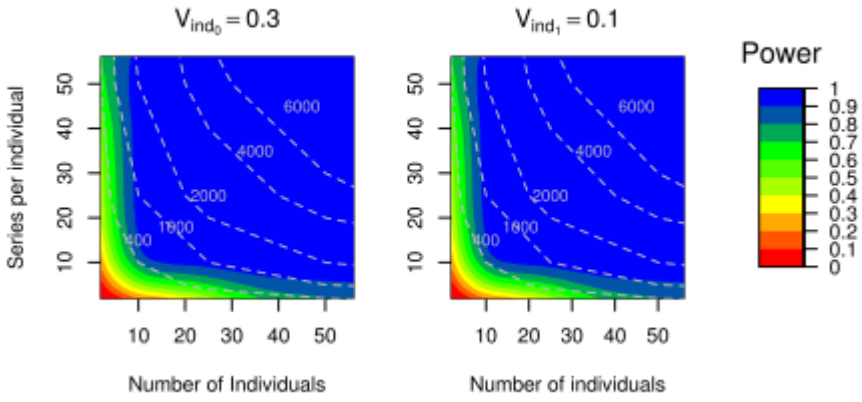

Figure 5. *Power to detect significant among-individual and among-series differences in slope. We measured power as the proportion of mixed models applied to the 100 simulated data sets that correctly rejected the null hypothesis of no variance at each of these levels, using a log-likelihood ratio.*

Simulation package

While we discuss optimal sampling designs for a specific type of situation in the Results section below, individual researchers may benefit from modelling a broader range of conditions applicable to their specific study system and sampling options. We therefore additionally developed a simulation package (MultiRR in R statistical environment (R Core Team 2014) that researchers can use to estimate bias and power of simulated data; this simulation package enables the inclusion of the unfortunate characteristics of real data, such as missing data, unbalanced observations across individuals, series, etc. The package may be used, first, to design an experimental study, and second, to assess whether the structure of an existing dataset allows reliable estimation of parameters with sufficient precision and statistical power.

A worked example

We illustrate how to implement our approach and quantify multi-level variation in reaction norms using a simulated example. Consider a hypothetical species (suitable for the type of **S**tatistical **QU**antification of **I**ndividual **D**ifferences proposed in this paper) of "squid" (acronym) that varies its anti-predator behavior in response to the level of predation risk. This mythological squid species varies the amount of ink it releases to avoid predators depending on the level of predation risk, but individuals become less responsive to variation in predation

risk with age (Fig. 2). Here, we are interested in estimating individual variation of ink release in response to different levels of perceived predation risk across four different age classes (1, 2, 3 and 4 months). Ink release was measured while an individual was exposed to either low or high simulated predation risk. Each individual was assayed twice in each risk context within each of the four ages. Observations belonging to a series were identified with a factor combining individual identity and age class (Fig. 6).

Figure 6

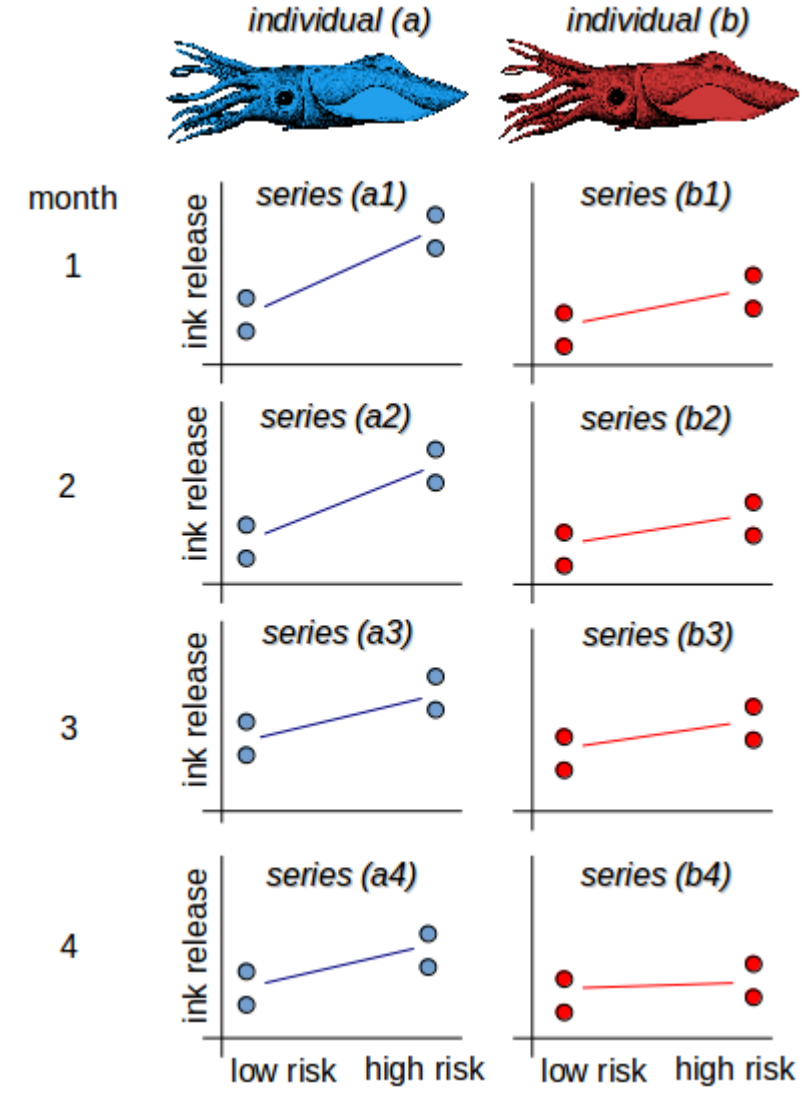

Figure 6. *Schematic representation of the experimental design to study multi-level variation in reaction norm components in a hypothetical species of squid. Anti-predator behavior for each individual was tested twice in two predation risk treatments (low vs. high predation risk), across 4 different life stages (1, 2, 3 and 4 months). We sampled 120 individuals, resulting in 480 series and 1920 observations. Series were identified with a factor combining individual identity and the month when the experiment was performed.*

We will use this worked example to show various issues regarding the estimation and interpretation of the parameter estimates derived from our proposed approach. We started with using the multiRR package to determine the optimal sampling design for this particular experiment. Doing so resulted in the following recommendation (Table S1): sample 120 individuals across the four ages (cf. 480 series and 1920 observations), as this design retrieves parameters with low bias and reasonable imprecision. We then proceeded with fitting a multi-level random regression model to a simulated data set with this level of replication (see Table 1 for details of the parameter settings used to simulate the data) in order to quantify variation in ink released by individual squids as a function of perceived predation risk. Predation risk level was fitted as a fixed effect covariate: 'low' was standardised to the value -0.5 and 'high' to the value 0.5, such that the intercept value was for the mean-centered environment (cf. Gelman 2008; Dingemanse & Dochtermann 2013). We fitted an additional interaction term between age (standardised to two standard deviation units) and predation risk to account for any population-average change in responsiveness across the different ages (see our discussion on adjusted slope repeatability above explaining why this may be a prudent decision). Random intercepts were included for individual and series; random slopes with respect to predation risk were also included at these two hierarchical levels. We assumed a Gaussian error distribution, fitted this model using the package lme4 (Bates et al. 2014), and simulated posterior distributions for parameter estimates using the sim function of the arm package (Gelman & Hill 2007). As predicted by our simulation study (see above and Table S1), the sampling design appeared appropriate since the parameter estimates were in correspondence with the true simulated values (compare "Simulation" with "Full model" in Table 1).

In order to demonstrate the characteristics and usefulness of our approach, we then proceeded to fit three more models and compared them to the full model. First, we fitted a model that lacked the interaction term between age and predation risk but was otherwise identical to the full model (Model 1, Table 1). We fitted this model to show that the presence of population-average multi-dimensional plasticity caused by an interaction between a within-individual within-series (predation risk) and within-individual among-series (age) factor results in among-series variation in plasticity if not accounted for. This was indeed the case (mode (95% credible intervals) for V_{series_1} was 0.27 (0.24, 0.30) for Model 1 compared to 0.19 (0.17, 0.21) for the Full model; Table 1). In other words, this comparison enabled us to

distinguish between within-individual among-series variation in reaction norm slopes due to identified (predation risk × age) versus unidentified population-average multidimensional plasticity. Slope repeatability estimates were, notably, logically different between the two models; Model 1 estimated a 'raw' repeatability of 0.39 which was naturally of lower magnitude than the (adjusted) repeatability of 0.53 estimated in the full model.

The second alternative model was constructed to investigate the consequences of fitting a classic single-level random regression (Model 2) instead of the proposed multi-level random regression (Full model) to datasets consisting of repeatedly expressed reaction norms. The aim of implementing model 2 (Table 1) was to show that when within-individual among-series variation in intercepts and slopes is not modelled, this variation will appear as unexplained residual variance. Indeed, the residual variance of Model 2 was about twice as high compared to the full (and other) models (Table 1).

The third alternative model was fitted to show that when among-individual variation in reaction norm parameters is not modelled (Model 3), this variation will be confounded with the among-series variation. This exercise implies that if variation in reaction norms would have been measured for just one series (cf. no repeated measures of reaction norms across series) then the among-individual variation in reaction norm components would have been conflated with among-series variation. Therefore, one is unable to formally investigate the occurrence of long-term individual differences in reaction norm parameters without collecting repeated series for the same set of individuals.

General Discussion

Evolutionary ecologists routinely estimate individual variation in phenotypic plasticity implemented as random regression mixed-effect models (equation 1). This paper details an expansion of this statistical framework to enable estimation of reaction norms that are repeatedly expressed within the same individual (Fig. 1). We also introduced the concept of repeatability of reaction norm components (intercept and slope), making it possible to empirically test adaptive theory predicting repeatable or unrepeatable plasticity under particular ecological conditions.

Table 1. Sources of variation in the amount of ink released by individual squid as a function of predation risk (two levels) and age class (four levels). We used univariate linear mixed-effect models with random intercepts and slopes (with respect predation risk; low risk coded as -0.5 and high risk as 0.5) at the level of the individual (*n* = 120 individuals) and series within individual (*n* = 480 series). Variances in intercepts are printed with subscript '0', variances in slopes with subscript '1' for among-individual ('ind'), among-series ('series'), and within series ('e') levels; intercept-slope covariances ('cov') are presented at each level. All values are reported as modes with 95% credible intervals.

	Simulation	Full model	Model 1	Model 2	Model 3
Fixed effects	β	β (95%CI)	β (95%CI)	β (95%CI)	β (95%CI)
Intercept	0.00	-0.06 (-0.16, 0.05)	-0.06 (-0.17, 0.05)	-0.06 (-0.17, 0.05)	-0.06 (-0.13, 0.01)
Slope	0.5	0.58 (0.48, 0.68)	0.58 (0.49, 0.68)	0.58 (0.48, 0.68)	0.58 (0.50, 0.65)
Age	0.00	0.01 (-0.08, 0.12)	-	0.01 (-0.05, 0.09)	0.01 (-0.13, 0.16)
Slope*age	-0.5	-0.48 (-0.61, -0.36)	-	-0.48 (-0.63, -0.34)	-0.48 (-0.62, -0.33)
Random effects	σ^2	σ^2 (95%CI)	σ^2 (95%CI)	σ^2 (95%CI)	σ^2 (95%CI)
Among individuals					
V_{ind_0}	0.3	0.31 (0.25, 0.38)	0.31 (0.25, 0.38)	0.36 (0.31, 0.451	-
V_{ind_1}	0.2	0.21 (0.17, 0.26)	0.17 (0.14, 0.21)	0.16 (0.13, 0.20)	-
Cov_{ind_0, ind_1}	0.1	0.13 (0.09, 0.17)	0.12 (0.08, 0.16)	0.17 (0.14, 0.21)	-
Within-individuals among-series					
V_{series_0}	0.3	0.34 (0.31, 0.38)	0.34 (0.30, 0.38)	-	0.62 (0.58, 0.67)
V_{series_1}	0.2	0.19 (0.17, 0.21)	0.27 (0.24, 0.30)	-	0.40 (0.36, 0.44)
$Cov_{series_0, series_1}$	0.1	0.12 (0.10, 0.15)	0.15 (0.12, 0.18)	-	0.29 (0.26, 0.33)
Residuals					
V_{e_0}	0.3	0.31 (0.29, 0.33)	0.31 (0.29, 0.33)	0.63 (0.59, 0.67)	0.30 (0.29, 0.33)
Repeatability	R	R (95%CI)	R (95%CI)	R (95%CI)	R (95%CI)
$R_{intercept}$	0.5	0.47 (0.45, 0.54)	0.48 (0.45, 0.53)	-	-
R_{slope}	0.5	0.53 (0.51, 0.59)	0.39 (0.37, 0.45)	-	-

We provide a set simulation tools to estimate bias, imprecision and power associated with different sampling schemes, and provide guidelines for the optimal sampling designs for studies aiming to estimate the repeatability of plasticity. Finally, we demonstrate using a simulation study how multi-dimensional plasticity can induce within-individual variation in reaction norm slopes and also how this proposed multi-level approach allows modelling patterns of variation that otherwise would not be revealed. The proposed multi-level random regression model thereby constitutes a useful method to test the adaptive theory predicting the ecological conditions favoring vs. disfavoring individual repeatability in phenotypic plasticity.

Revealing biology in residual within-individual variance

Most statistical descriptions of labile phenotypic characters are characterised by substantial amounts of residual within-individual variance, and there is growing awareness that this often overlooked residual variance may contain important biological information (Cleasby & Nakagawa 2011; Westneat et al. 2014b). The statistical framework proposed here takes up the challenge to start explaining residual within-individual variance by explicitly acknowledging that reaction norms can vary both among and within individuals. This approach will allow formal testing of hypotheses for adaptive individual differences in labile phenotypic characters that make explicit predictions about the presence of among- *and* within-individual variation in reaction norm components. For example, theoretical analyses predict that individual differences in plasticity can be favored when the payoffs for plasticity are negatively frequency-dependent (Wolf et al. 2008). However, frequency-dependence alone makes no predictions regarding the extent to which individuals should be consistent in their reaction norm components. Consistency in reaction norm components is only predicted when the payoffs for plasticity are state-dependent (e.g. an individual's experience with plasticity reduces the costs of future expressions of plasticity; Wolf et al. 2008). Thus, lack of individual repeatability of reaction norm components would imply that individual differences in plasticity are not state-dependent, or that the relevant state variable was not stable over the time scale represented by each series. This latter explanation implies that reaction norms with respect to a single environmental axis can vary between series of the same individual as a function of other (unmeasured) environmental axes. Such multi-dimensionality of reaction norms represents an important but largely overlooked biological phenomenon (Westneat et al. 2009, 2011, 2014a); the presence of variance in reaction norm components among series of

the same individual thus signals that key environmental components contributing to multi-dimensional reaction norms were overlooked.

The approach outlined here will also facilitate more explicit consideration of the ecological factors that promote stability of individual phenotypes. For example, individual predictability of behaviour should be favored for social behaviours when the outcomes of interactions can be observed by others (Dall et al. 2004). By being predictable (i.e. consistent in both intercept and slope of a reaction norm), individuals may be able to avoid lengthy (costly) interactions with others, for example if they only escalate an aggressive interaction when the costs of not behaving aggressively are high (Dall et al. 2004).

Several studies have also recently begun to explore among-individual differences in residual stability, where some individuals are more stable (‘predictable’) compared to others when major environmental gradients were assumed to be controlled for in the statistical analysis (Stamps et al. 2012; Briffa et al. 2013; Biro & Adriaenssens 2013; Cleasby et al. 2014). As detailed in this paper, such repeatability of residual within-individual variation, verbally also called ‘individual differences in intra-individual variability’ (Stamps et al. 2012), might occur not because individuals generally differ in predictability *per se* (Briffa et al. 2013) but rather because repeatable individual variation in phenotypic plasticity was not captured in the statistical model. In other words, the finding of among-individual variation in within-individual residual variance should perhaps best be taken as a starting point to investigate whether individuals genuinely differed in predictability versus whether they appeared to differ because the fitted statistical model was, in fact, incomplete (sensu Westneat et al. 2014b).

Multi-level analyses and the estimation of different types of repeatability

Repeatability has routinely been applied to estimate the proportion of total phenotypic variance in a dataset that is attributable to differences among individuals (Nakagawa & Schielzeth 2010). Repeatability (R) is often quantified as the proportion of phenotypic variance not attributable to fixed effects explained by differences between individuals (called 'adjusted' repeatability; (Nakagawa & Schielzeth 2010) and may readily be calculated from the multi-level random regression model (Eqn. 2) as (Equation 5):

$$R = \frac{V_{ind_0}}{V_{ind_0} + V_{series_0} + V_{e_0}} \quad \text{(Eqn. 5)}$$

This classic repeatability differs distinctly from repeatability of average behaviour (reaction norm intercept repeatability ($R_{intercept}$); Eqn. 3a) because the latter proportion does not include the residual variance (V_{e_0}) in its denominator. Differences among individuals at any time point are due to permanent environmental effects and genetic differences (causing V_{ind_0}) but also due to among-individual differences in environmental conditions causing short term consistency (causing). V_{series_0} The ratio derived from equation 5 (R), represents the amount of the total phenotypic variation attributed to permanent environmental effects and genetic differences. In contrast, by excluding the environmental variation that is not causing short term consistency (V_{e_0}) from the denominator, the ratio derived from equation 3a ($R_{intercept}$) will represent the amount of differences between-individuals due to long term consistency.

Recent meta-analysis of behavior has shown that repeatability estimates are higher when repeated measures are taken closely in time (Bell et al. 2009); this finding has stimulated interest in the statistical estimation of short- versus long-term repeatability (Boulton et al. 2014). That repeatability should vary with the time interval between repeated measures makes intuitive sense because environmental conditions affecting the phenotype often show temporal autocorrelation and thus cause short-term individual consistency in a dataset. The mixed-effect modelling approach proposed in this paper also represents an ideal tool for estimating short- versus long-term repeatability, and thereby answers the call for the development of tools to estimate repeatability at different time scales (Bell et al. 2009; Boulton et al. 2014). Specifically, short-term repeatability in the average environment is calculated from the multi-level random regression model (Eqn. 2) as (Equation 6):

$$R_{short\ term} = \frac{V_{ind_0} + V_{series_0}}{V_{ind_0} + V_{series_0} + V_{e_0}} \quad \text{(Eqn 6)}$$

Equation 5 in contrast represents long-term repeatability. The distinction between the two equations is that short-term repeatability includes variance among series (i.e. short-term consistency caused by among-individual differences in environmental conditions) in the

numerator, which causes short-term repeatability to be equal to or higher than long-term repeatability. In our worked example, short term repeatability was, as expected, higher than long term repeatability (0.67 vs. 0.32). This usage of extra random effects capturing the temporal dependency of observations, such as our 'series', may also be applied to the partitioning of individual intercepts across multiple time scales. For example, nesting the unique combination of individual and day ('day series') within the unique combination of individual and month ('month series') within the unique combination of individual in year ('year series)' within the individual would enable a detailed partitioning of repeatability across days, months, and years, respectively.

Bias

An important question in statistical analyses of multi-level data is whether the estimates retrieved from the model are unbiased. Bias, or inaccuracy, reflects the level of disagreement between the model's estimate and the 'true' value (e.g. van de Pol 2012). Applying this issue to multi-level random regression analyses, we showed that the optimal sampling scheme will depend on the parameter(s) of interest, and what aspect of the models researchers want to optimize (bias, imprecision or power). We also acknowledge that the amount of data necessary to properly estimate multi-level random regression parameters is very specific to study systems, and this is why we created a set of easy to use functions to allow researchers tailoring their simulation studies to the specifics of their systems.

Our simulations were used to determine bias caused by insufficient replication (i.e. in terms of number of individuals or series), but bias may also occur for various other reasons. For example, individuals might not all be sampled over the same range of the environmental gradient fitted as the covariate in the random regression model. Such 'repeatability' of environmental conditions during sampling may introduce major biases in the estimation of both fixed and random parameters (van de Pol 2012), though statistical techniques such as 'within-subject centering' (Dingemanse & Dochtermann 2013) may be applied to alleviate such concerns (but see Phillimore et al. 2010). Our simulation and worked example, explicitly considered ideal, balanced, data sets where such concerns do not apply but our simulation package does enable researchers to address such problems themselves.

Our proposed hierarchical structuring of random effects will reduce the bias that may occur because the specified mixed-effect model is incomplete (Westneat et al. 2014b), which

is important even if the aim of a particular analysis is not to quantify multi-level variation in reaction norm components. Bias due to incomplete models would occur, for example, in cases where within-individual residual variances are heterogeneous in nature but homogeneous errors are instead assumed (Cleasby & Nakagawa 2011; Westneat et al. 2014b). This problem is illustrated by a study on great tits where the amount of variation in phenotypic plasticity (with respect to how conspecific density affected clutch size) was biased upwards in models that failed to consider heterogeneous within-individual among-year variance (Nicolaus et al. 2013). Failure to model random slopes, or intercept-slope covariance, would, similarly, lead to biased estimates of fixed and random effects (Schielzeth & Forstmeier 2009; Barr et al. 2013). While intercept slope variation and covariance is routinely modelled at the individual level, this is certainly not the case for the within-individual level because repeatedly expressed reaction norms have rarely been modelled (Dingemanse & Wolf 2013b; this paper). Our proposed expansion of the mixed-effect model thus helps capturing such heterogeneity, acknowledges the temporal dependencies between observations within the same series and helps avoiding incomplete models and biased estimates.

In summary, our multi-level implementation of random slope models acknowledges that patterns of heterogeneous variance can exist at multiple hierarchical levels. While there is increasing awareness that residual within-individual variance might often be heterogeneous in the data set as a whole (Cleasby & Nakagawa 2012), this paper recognizes that assumptions regarding homogeneity of variances do not just apply to the residual level but may vary with respect to specific fixed effects (Westneat et al. 2012), random effects (Nicolaus et al. 2013), or interactions between random effects such as population differences in among-individual variance (Westneat et al. 2014a) or individual differences in within-individual variance (Stamps et al. 2012; Briffa et al. 2013; Biro & Adriaenssens 2013).

Conclusions

This paper introduces a statistical framework that enables the estimation of among- and within-individual variation in reaction norms. Multi-level variation in reaction norms characterise a multitude of labile phenotypic characters such as those that respond to multiple environmental stimuli (multi-dimensional reaction norms). The proposed framework enables the calculation of new standardised indices, namely the repeatability of an average phenotype

(reaction norm intercept) and the repeatability of level of responsiveness (reaction norm slope), which may readily be compared across studies. The proposed methodology also enables the calculation of short- versus long-term repeatability, and contributes to our understanding of biological variation that may otherwise be dubbed residual within-individual variance.

Acknowledgements

We thank the members of the SQUID working group (Hassen Allegue, Ned Dochtermann, Lazlo Garamszegi, Shinichi Nakagawa, Denis Réale, Holger Schielzeth, and Dave Westneat) and also to Julien Martin for feedback and discussion on the ideas presented in this paper. The study was supported by the Max Planck Society, YA-A was supported by the German Academic Exchange Service (DAAD) and the International Max Planck Research School for Organismal Biology, and KJM was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). The study was conceived by NJD and YA-A; YA-A conducted the simulations and statistical analyses; the manuscript was prepared by NJD and YA-A, and co-written by KJM.

References

Araya-Ajoy, Y. G. & Dingemanse, N. J. 2014. Characterizing behavioural “characters”: an evolutionary framework. *Proceedings. Biological sciences / The Royal Society*, **281**, 20132645.

Auld, J. R., Agrawal, A. a & Relyea, R. a. 2010. Re-evaluating the costs and limits of adaptive phenotypic plasticity. *Proceedings. Biological sciences / The Royal Society*, **277**, 503–11.

Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, **68**, 255–278.

Bates, D., Mächler, M. & Bolker, B. 2014. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*,

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. & Adriaenssens, B. 2013. Predictability as a Personality Trait: Consistent Differences in Intraindividual Behavioral Variation. *The American Naturalist*, **182**, 621–629.

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

Briffa, M., Rundle, S. D. & Fryer, A. 2008. Comparing the strength of behavioural plasticity and consistency across situations: animal personalities in the hermit crab Pagurus bernhardus. *Proceedings. Biological sciences / The Royal Society*, **275**, 1305–11.

Briffa, M., Bridger, D. & Biro, P. A. 2013. How does temperature affect behaviour? Multilevel analysis of plasticity, personality and predictability in hermit crabs. *Animal Behaviour*, **86**, 47–54.

Brommer, J. E., Rattiste, K. & Wilson, A. J. 2008. Exploring plasticity in the wild: laying date-temperature reaction norms in the common gull Larus canus. *Proceedings of the royal society of london.*, **275**, 687–93.

Cleasby, I. R. & Nakagawa, S. 2011. Neglected biological patterns in the residuals. *Behavioral Ecology and Sociobiology*, 2361–2372.

Cleasby, I. & Nakagawa, S. 2012. The influence of male age on within-pair and extra-pair paternity in passerines. *Ibis*, **154**, 318–324.

Cleasby, I. R., Nakagawa, S. & Schielzeth, H. 2014. Quantifying the predictability of behaviour: statistical approaches for the study of between-individual variation in the within-individual variance. *Methods in Ecology and Evolution*, n/a–n/a.

Coppens, C. M., de Boer, S. F. & Koolhaas, J. M. 2010. Coping styles and behavioural flexibility: towards underlying mechanisms. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **365**, 4021–8.

Dall, S. R. X., Houston, A. I. & McNamara, J. M. 2004. The behavioural ecology of personality: consistent individual differences from an adaptive perspective. *Ecology Letters*, **7**, 734–739.

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

Dingemanse, N. J. & Wolf, M. 2013. Between-individual differences in behavioural plasticity within populations: causes and consequences. *Animal Behaviour*, **in press**, 1–9.

Dingemanse, N. J., Edelaar, P. & Kempenaers, B. 2010a. Why is there variation in baseline glucocorticoid levels? *Trends in Ecology and Evolution*, **25**, 261–262.

Dingemanse, N. J., Kazem, A. J. N., Réale, D. & Wright, J. 2010b. Behavioural reaction

norms: animal personality meets individual plasticity. *Trends in ecology & evolution*, **25**, 81–9.

Enders, C. K. & Tofighi, D. 2007. Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychological methods*, **12**, 121–38.

Forsman, a. 2014. Rethinking phenotypic plasticity and its consequences for individuals, populations and species. *Heredity*, 1–9.

Gelman, A. 2008. Scaling regression inputs by dividing by two standard deviations. *Statistics in medicine*, 2865–2873.

Gelman, A. & Hill, J. 2007. *Data analysis using regression and multilevel/hierarchical models.* New York: Cambridge University Press.

Jenkins, S. H. 2011. Sex differences in repeatability of food-hoarding behaviour of kangaroo rats. *Animal Behaviour*, **81**, 1155–1162.

Koolhaas, J. M., de Boer, S. F., Coppens, C. M. & Buwalda, B. 2010. Neuroendocrinology of coping styles: Towards understanding the biology of individual variation. *Frontiers in Neuroendocrinology*, **31**, 307–321.

Lema, S. C. & Kitano, J. 2013. Hormones and phenotypic plasticity□: Implications for the evolution of integrated adaptive phenotypes Endocrine Systems as Multidimen- sional Regulatory Networks. *Current zoology*, **59**, 506–525.

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

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

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

McNamara, J. M., Stephens, P. a, Dall, S. R. X. & Houston, A. I. 2009. Evolution of trust and trustworthiness: social awareness favours personality differences. *Proceedings. Biological sciences / The Royal Society*, **276**, 605–13.

Nakagawa, S. & Schielzeth, H. 2010. Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biological reviews of the Cambridge Philosophical Society*, **85**, 935–56.

Nicolaus, M., Brommer, J. E., Ubels, R., Tinbergen, J. M. & Dingemanse, N. J. 2013. Exploring patterns of variation in clutch size–density reaction norms in a wild passerine bird. *Journal of Evolutionary Biology*, **26**, 2031–2043.

Nussey, D. H., Kruuk, L. E. B., Donald, A., Fowlie, M. & Clutton-Brock, T. H. 2006. The rate of senescence in maternal performance increases with early-life fecundity in red deer. *Ecology Letters*, **9**, 1342–1350.

Nussey, D. H., Wilson, a J. & Brommer, J. E. 2007. The evolutionary ecology of individual phenotypic plasticity in wild populations. *Journal of Evolutionary Biology*, **20**, 831–44.

Phillimore, A. B., Hadfield, J. D., Jones, O. R. & Smithers, R. J. 2010. Differences in spawning date between populations of common frog reveal local adaptation. *Proceedings of the National Academy of Sciences of the United States of America - PNAS*, **107**, 8292–8297.

Piersma, T. & Drent, J. 2003. Phenotypic flexibility and the evolution of organismal design. *Trends in Ecology & Evolution*, **18**, 228–233.

R Core Team. 2014. R: A language and environment for statistical computing.

Schaeffer, L. R. 2004. Application of random regression models in animal breeding. *Livestock Production Science*, **86**, 35–45.

Schielzeth, H. & Forstmeier, W. 2009. Conclusions beyond support: overconfident estimates in mixed models. *Behavioral Ecology*, **20**, 416–420.

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

Schlichting, C. & Pigliucci, M. 1998. *Phenotypic Evolution: A Reaction Norm Perspective.* Sinauer.

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

Van de Pol, M. 2012. Quantifying individual variation in reaction norms: how study design affects the accuracy, precision and power of random regression models. *Methods in Ecology and Evolution*, **3**, 268–280.

Westneat, D. F., Stewart, I. R. K. & Hatch, M. I. 2009. Complex interactions among temporal variables affect the plasticity of clutch size in a multi-brooded bird. *Ecology*, **90**, 1162–1174.

Westneat, D. F., Hatch, M. I., Wetzel, D. P. & Ensminger, A. L. 2011. Individual variation in parental care reaction norms: integration of personality and plasticity. *The American Naturalist*, **178**, 652–67.

Westneat, D. F., Schofield, M. & Wright, J. 2012. Parental behavior exhibits among-individual variance, plasticity, and heterogeneous residual variance. *Behavioral Ecology*, **24**, 598–604.

Westneat, D. F., Bókony, V., Burke, T., Chastel, O., Jensen, H., Kvalnes, T., Lendvai, Á. Z., Liker, A., Mock, D., Schroeder, J., Schwagmeyer, P. L., Sorci, G. & Stewart, I. R. K.

2014a. Multiple aspects of plasticity in clutch size vary among populations of a globally distributed songbird. *Journal of Animal Ecology*, **83**, 876–887.

Westneat, D. F., Wright, J. & Dingemanse, N. J. 2014b. The biology hidden inside residual within-individual phenotypic variation. *Biological Reviews*, n/a–n/a.

Wolf, M., van Doorn, G. S. & Weissing, F. J. 2008. Evolutionary emergence of responsive and unresponsive personalities. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 15825–30.

Wright, J., McDonald, P. G., te Marvelde, L., Kazem, A. J. N. & Bishop, C. M. 2010. Helping effort increases with relatedness in bell miners, but “unrelated” helpers of both sexes still provide substantial care. *Proceedings. Biological sciences / The Royal Society*, **277**, 437–45.

Chapter 3

Interacting personalities: Behavioural Ecology meets Quantitative Genetics

Niels J. Dingemanse & Yimen G. Araya-Ajoy

Trends in Ecology and Evolution (2015), 282, 88-97

CellPress

Interacting personalities: behavioural ecology meets quantitative genetics

Niels J. Dingemanse[1,2] and Yimen G. Araya-Ajoy[1]

[1] Evolutionary Ecology of Variation Research Group, Max Planck Institute for Ornithology, Seewiesen, Germany
[2] Behavioural Ecology, Department of Biology, Ludwig Maximilians University of Munich, Planegg-Martinsried, Germany

Behavioural ecologists increasingly study behavioural variation within and among individuals in conjunction, thereby integrating research on phenotypic plasticity and animal personality within a single adaptive framework. Interactions between individuals (cf. social environments) constitute a major causative factor of behavioural variation at both of these hierarchical levels. Social interactions give rise to complex 'interactive phenotypes' and group-level emergent properties. This type of phenotype has intriguing evolutionary implications, warranting a cohesive framework for its study. We detail here how a reaction-norm framework might be applied to usefully integrate social environment theory developed in behavioural ecology and quantitative genetics. The proposed emergent framework facilitates firm integration of social environments in adaptive research on phenotypic characters that vary within and among individuals.

Personality, plasticity, and social interactions

Behavioural ecology research increasingly acknowledges the characteristic multilevel nature of animal behaviour [1], investigating within-individual (cf. phenotypic plasticity) and among-individual variation (cf. animal personality) in conjunction [2] (see Glossary). Adaptive explanations for behavioural variation centre upon the proposition that 'state' (features of organisms affecting the balance of costs and benefits of behavioural actions [3]) varies both within and among individuals, explaining behavioural variation at both levels [3–6]. Adaptive personality theory, for example, explains among-individual variation in behaviour as an adaptation to endogenous features of individuals [4,5], such as metabolism [7] and cognitive ability [8]. Exogenous features, particularly social environments, have more recently come to the foreground as key state variables shaping variation among individuals [9–11]. Social environments are of major importance because interactions between conspecifics impose a diverse array of selective pressures on various behaviours.

Models of adaptive behaviour imply a key role for social interactions [11]. Classic examples such as hawk–dove, producer–scrounger, and leader–follower games demonstrate how interactions often induce selection favouring behavioural variation [12]. Interactions can give rise to either adaptive within-individual variation (cf. plastic, conditional strategies) or adaptive among-individual variation (cf. alternative, fixed strategies) [6,11]. Adaptive theory, for example, implies that predictability in aggressiveness can be favoured when it allows interacting individuals to avoid costly fights [13]. The resulting among-individual variation has been suggested to favour the emergence of 'socially responsive' [13] individuals who adjust their behaviour as a function of the previous

Glossary

Among-individual variation: individual differences in average phenotype across multiple observations.
Animal personality: among-individual variation in behaviour attributable to the combined influences of genetic effects and environmental effects that permanently affect the phenotype of an individual [2,6]. Pseudo-personality occurs when estimates of personality are inflated because of individual repeatability in environmental conditions that cause nonpermanent effects on behaviour [31,36].
Direct genetic effect (DGE): allelic variation in genes affecting the phenotype, where the phenotype of an individual is directly affected by its own genes [21].
Emergent character: a phenotypic character representing a characteristic or an outcome of an interaction rather than of an individual, such as the duration or intensity of a fight [43].
Indirect genetic effect (IGE): environmental influences on the phenotype of an individual resulting from the expression of genes in another conspecific [17,21].
Interactive phenotype: a phenotypic characteristic of an individual whose expression is affected by the phenotype of (a) conspecific(s).
Phenotypic gambit: an approach to the study of behavioural adaptation [33] viewing natural selection as an optimising process that is ultimately unconstrained by genetic architecture [59].
Reaction norm (RN): set of phenotypes that a genotype or individual produces as a function of an environmental gradient. Throughout this paper, we focus on individual-level reaction norms [2,34].
Social environment: environmental component of the phenotype caused by interactions with conspecifics.
Social responsiveness: phenotypic plasticity in response to the phenotype expressed by a conspecific, estimated as the slope of an individual-level reaction norm. Socially responsive individuals are characterised by a nonzero interaction coefficient (Ψ).
Trait-based approach: a statistical approach where phenotypes of focals are represented as a function of the phenotypic characteristics of conspecifics [53]. This dependency is captured by an interaction coefficient (Ψ).
Variance-partitioning approach: a statistical approach where phenotypic variance is partitioned in variance attributable to different effects [53]. Variance in phenotype of focals might, for example, be decomposed into variance explained by the identity of the focal versus social partner, or into variance explained by direct genetic effects versus indirect genetic effects [53].
Within-individual variation: phenotypes vary within individuals across instances, caused by nonpermanent environmental effects on the phenotype of an individual [2,6,31,34]. Throughout, we assume that variance attributable to measurement error represents a negligible component of within-individual variance.

Corresponding author: Dingemanse, N.J. (ndingemanse@orn.mpg.de).
Keywords: reaction norm; indirect genetic effects; interacting phenotypes; personality; quantitative genetics; phenotypic plasticity; social environments; social evolution.

0169-5347/
 http://dx.doi.org/10.1016/j.tree.2014.12.002

interactions of their social partner (cf. within-individual variance resulting from adaptive phenotypic plasticity) [14], which in turn causes intensified selection favouring further individual differentiation in various types of behaviour (e.g., aggressiveness [15], cooperation [14], or coordination [16]). Similarly, repeated interactions between individuals cooperating in stable social groups have been proposed to increase among-individual (but decrease within-individual) variation in behaviour [10] because negative frequency-dependent selection favours division of labour among individuals (cf. social niche specialisation [9]). Thus, social interactions might give rise to personality, plasticity, and individual differences in social responsiveness [9,11,13].

Social environment effects in quantitative genetics

Quantitative geneticists have studied social environments from a different perspective. Their emphasis has been on predicting evolutionary responses to selection [17,18]. Quantitative genetic theory developed by animal breeders and evolutionary biologists implies that social environments can have major evolutionary repercussions when heritable phenotypes affect the phenotypes of other conspecific individuals [19,20]. In such cases, the social environment is itself heritable because of 'indirect genetic effects' (IGEs) and, thus, is evolvable [17,21]. IGEs represent a special form of phenotypic plasticity where environmental effects on the phenotype of an individual are caused by the expression of genes in another conspecific [21]; the familiar 'direct genetic effects' (DGEs) instead occur when the phenotype of an individual is directly affected by its own genes. Genetic variation in maternal investment influencing offspring development (cf. maternal genetic effects [22]) and genetic characteristics of social partners affecting life-history decisions of mates [23] represent examples of IGEs. Importantly, IGEs influence evolutionary responses to selection, such as when there are functional interactions between traits of interacting individuals [24] or when DGEs and IGEs are genetically correlated [21]. In gulls, for instance, genes expressed in females contributing to early laying (DGEs) are negatively correlated with genes expressed in males facilitating early laying in female partners (IGEs) [23]. Such sexually antagonistic effects can impose constraints on evolution [25]. Positive genetic correlations might instead speed up evolutionary responses (depending on the selective landscape [17,20]); genes for aggressiveness in mice (a DGE), for example, correlate positively with genes eliciting aggressiveness in opponents (an IGE) [26]. Thus, phenotypic plasticity as a function of phenotypes expressed by conspecifics (i.e., social responsiveness) represents a key factor in the evolutionary process. However, little is known about the ecological conditions (dis)favouring indirect genetic effects [27] and whether social responsiveness is heritable and evolvable [28,29].

Behavioural ecology meets quantitative genetics

In this opinion article, we propose a reaction-norm framework to combine social environment theory developed in behavioural ecology and quantitative genetics, and to facilitate cross-fertilisation between these research fields. We detail how quantitative genetics approaches might be usefully incorporated in behavioural ecology research (cf. [10,30,31]) to empirically study the adaptive nature of 'social responsiveness'. Conversely, we argue that behavioural ecology theory on this topic usefully provides quantitative genetics with predictions concerning ecological conditions (dis)favouring the evolution of variance components such as indirect effects (cf. [32]). Behavioural ecologists apply a 'phenotypic gambit' [33] in their adaptive studies; in this opinion article, we adopt this approach by focusing on among-individual (rather than additive genetic) variation; both approaches intricately contribute to our understanding of evolutionary processes (Box 1). The proposed framework enables integration of social environments between distinct fields of evolutionary biology.

Phenotypes as environmental gradients

Incorporating social environments into studies of personality and plasticity requires a particular way of thinking about both behaviour and social environments. Instead of characterising individuals by their behaviour, we view the

Box 1. Behavioural ecology, variance components, and evolutionary adaptation

Quantitative genetics focuses on predicting evolutionary responses to selection, and this explicitly requires the partitioning of phenotypic variation in traits (and fitness) in genetic versus environmental components [22]. Behavioural ecology, by contrast, commonly applies a 'phenotypic gambit' [33], viewing natural selection as an optimising process that is ultimately unconstrained by genetic architecture [59], which might therefore be studied at the phenotypic level. Behavioural ecology approaches nevertheless contribute importantly to our understanding of evolutionary processes. Specifically, interest in (repeatable) among-individual differences has stimulated the development of theory predicting the ecological conditions under which natural (and sexual) selection (dis)favour specific (co)variance components [31,32] such as among-individual (co)variance [4,5,13], within-individual variance [60], and among-individual variation in behavioural plasticity [6,11,14]. Here, behavioural ecology theory implies a key role for ecology (cf. resource availability, predation risk [61]) in causing selection (dis)favouring among-individual variance [4,5]. Models typically involve adaptive state-dependence of behaviour [3–5,13,61], leading to testable predictions concerning the magnitude of permanent-environmental (e.g., [62]) and within-individual variances (e.g., [63]). The expression of such non-genetic variance components directly affects the heritability of phenotypic characters, hence their evolutionary potential [22]. Despite its application of a phenotypic gambit, behavioural ecological theory therefore contributes substantially to our understanding of evolutionary processes. At the same time, their focus on 'unpartitioned' among-individual variance hampers the application of quantitative genetics theory in predicting evolutionary responses to selection.

Adaptive theory concerning the emergence of direct (cf. among-individual) and indirect (cf. social partner) effects developed by behavioural ecologists, importantly, does not hinge upon the nature of state-dependence: various types of predictive theory apply generally to both heritable (cf. additive genetic) and nonheritable (cf. permanent environmental) parts of among-individual variance components [4]. In other words, behavioural ecology theory concerning the emergence of social responsiveness (cf. indirect effects), or among-individual variation in social responsiveness, can readily be utilised to study the ecological conditions (dis)favouring both indirect environmental effects and indirect genetic effects, and thereby meaningfully enables the integration of ecology into the study of heritable variation.

function describing the dependency of behaviour on the environment as a key individual character [2,34]. Such a 'reaction norm' (RN) integrates information about its average behaviour (cf. personality) and level of phenotypic plasticity in response to environmental gradient(s) [2,34], specifying the specific (linear or nonlinear) function between response (behaviour) and predictor (environmental) variable(s) [2,34]. RNs of individuals are estimable whenever the phenotype of an individual is assayed over a range of conditions [2,31], such as its vigilance under different levels of predation risk [35].

The simplest form of individual RN summarises observations (instances i) of a behaviour (P_{ij}) of an individual (j), and environmental state (x_{ij}; typically a mean-centred covariate [31]), into a linear function that specifies its RN elevation (I_j; its behaviour in the average environment) and slope (β_j; its level of responsiveness), and the deviation of each instance from its estimated RN (e_{ij}) (i.e., $P_{ij} = I_j + \beta_j x_{ij} + e_{ij}$). We propose incorporation of social environments into this framework by letting the environmental covariate (x) represent the phenotype of an interaction partner (Figure 1). Such RNs represent characteristics of individuals that are quantifiable whenever the social environment (cf. phenotype of conspecifics) varies across behavioural observations of an individual (Figure 1). This type of RN has the particularity that within-individual variation in social environments can be the result of two interacting processes. First, the social partner(s) of an individual change(s) phenotype. Second, individuals switch social partners, leading to a change in social environment either because partners show repeatable differences in phenotype (e.g., gender, size, or personality) or because partner behaviour is plastic and changes between instances. Adjustments in parental care in response to moment-to-moment changes in prey delivery of partners in birds [36] exemplify variation in social environments caused by within-partner variation in phenotype. Changes in aggressiveness as a function of partner size in fish [37] exemplifies variation in social environments caused by among-partner variation in phenotype (see Table 1 for further examples).

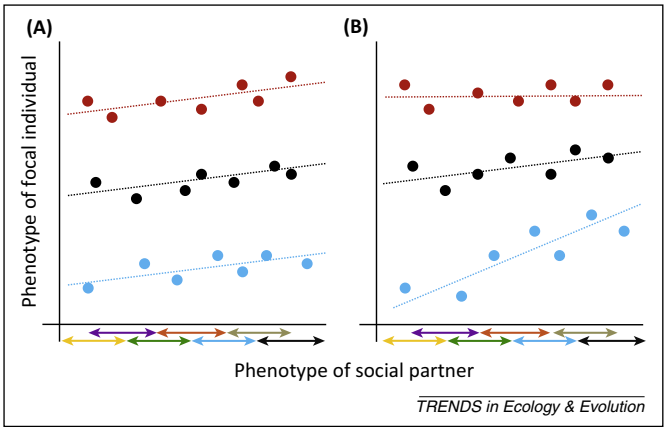

Figure 1. Reaction-norm (RN) plots, where the phenotype (dots) expressed by a focal individual is plotted as a function (lines) of the phenotype expressed by an interaction partner. Each panel shows a RN plot for each of three individuals (colours) each assayed seven times. Socially responsive (cf. plastic) versus unresponsive (cf. nonplastic) individuals have nonhorizontal versus horizontal RNs, respectively. The seven coloured double-headed arrows (printed below each x axis) represent seven individual interaction partners; each partner is phenotypically plastic, with the length of its arrow representing its phenotypic range, but interaction partners also differ in average phenotype (cf. 'personality' for a behavioural phenotype). Consequently, variation in social environments results from the combined influences of among- and within-individual variation in phenotypes expressed by interaction partners. For simplicity, both panels depict scenarios quantifying general responsiveness to the phenotypes of interaction partners (Ψ); for example, each focal was assayed once with each social partner. Responsiveness to the repeatable (Ψ_A; A for among) and plastic (Ψ_W; W for within) part of the phenotype of the partner could be incorporated provided that the phenotype of each partner had been quantified repeatedly (Box 3). This type of RN can vary among individuals, in terms of elevation **(A,B)** and slope (B), and these two RN components can be correlated (B). Individuals in (A) are all socially responsive ($\Psi \neq 0$) but do not differ in responsiveness. By contrast, there is among-individual variation in responsiveness in (B): some individuals are socially responsive (black, blue) but others are not (red), representing 'reactive' and 'proactive' individuals, respectively [40] (for details see main text).

Social responsiveness and pseudo-personality

The simple presence of behavioural responsiveness to the phenotype of social partners (cf. $\beta \neq 0$; Figure 1A) can lead to the appearance of personality where none exists (i.e., 'pseudo-personality' [31,36]). With regard to social environment effects, this occurs when animals respond plastically to phenotypes that are repeatable in their partners, while simultaneously showing repeatability in partner identity (cf. stable social pairs or groups; Box 2). Among-individual variation in phenotypes of partners would then automatically create among-individual variation in the behaviour of focal individuals. Such apparent personality differences would disappear if partner identity or phenotype was fitted as a predictor in statistical analyses of RNs (Box 2). By contrast, social environments can also result in genuine among-individual differentiation in behaviour (cf. among-individual variance in RN elevation). Such long-term effects are well known from mammals, where the parenting style of the mother influences a variety of behaviours (e.g., anxiety) that her offspring express in adulthood [38].

Individuality in social responsiveness

Viewing RNs as individual characteristics [2,34] implies that they might vary across populations or individuals, whether in elevation (cf. personality) or slope (cf. responsiveness). This notion is confirmed for RNs in general [2,34,39], but in ecology not widely studied in the context of responsiveness to phenotypes of conspecifics (Table 1). Research on behavioural stress physiology forms an exception where individuality in responsiveness to environmental stimuli is explicitly recognised [40]. Here, 'reactive' animals modify aggressive behaviour as a function of aggressive intent signalled by interaction partners, whereas 'proactive' individuals instead behave aggressively irrespective of social context (e.g., [41]). Thus, elevations and slopes of RNs quantifying responsiveness to conspecific phenotypes can vary and correlate across individuals [2,34] (note, however, that the magnitude and sign of such estimates depend on scaling and centring [42]). Proactive animals, for example, have relatively high RN elevations in combination with shallow RN slopes (red individual in Figure 1B), whereas reactive individuals have low RN elevations in combination with steeper RN slopes (blue individual in Figure 1B). Adaptive theory on social personalities explicitly predicts such covariance between personality and social responsiveness [11,14–16].

Table 1. Examples of social environment effects within and among individuals drawn from the ecological literature to which adaptive concepts detailed in this opinion article could be applied

Species	Method[a]	Phenotype of partner → phenotype of focal[b]	Interaction partner(s)[c]	Variation in Ψ[d]	Ref.
Insects					
Pacific field cricket (*Teleogryllus oceanicus*)	TB	Song (a) → female preference (w)	Different	Yes	[28]
Fruit fly (*Drosophila serrata*)	TB	Unknown (a) → cuticular hydrocarbons (w)	Different	Yes	[29]
Fish					
Blackfin pearlfish (*Austrolebias nigripinnis*)	VP	Unknown (a) → egg size (w)	Different	No	[64]
Green swordtail (*Xiphophorus helleri*)	VP/TB	Size (a) → aggression (w)	Different	No	[37]
Trinidadian guppy (*Poecilia retiuclata*)	TB	Male ornament (a) → female choice (w)	Different	Yes	[65]
Reptiles					
Side-blotched lizard (*Uta stansburiana*)	TB	Maternal investment (w) → escape behaviour (a)	Same	No	[66]
Birds					
Common gull (*Larus canus*)	VP	Unknown (a) → lay-date (w)	Same	No	[23]
House sparrow (*Passer domesticus*)	TB	Provisioning (w) ↔ provisioning (w)	Same	No	[36]
Great tit (*Parus major*)	TB	Begging (w) ↔ provisioning (w)	Same	No	[67]
	VP	Unknown (a) ↔ nestling mass (w)	Different	No	[49]
	TB	Exploration (a) → promiscuity (w)	Different	No	[68]
Japanese quail (*Coturnix japonica*)	TB	Hormone levels (a) → boldness (w)	Same	No	[66]
Mammals					
Deer mouse (*Peromyscus maniculatus*)	VP	Unknown (a) ↔ aggression (w)	Different	No	[26]
Human (*Homo sapiens*)	TB	Maternal care (w) → stress reactivity (a)	Same	No	[38]
House mouse (*Mus musculus*)	TB	Aggression (w) ↔ aggression (w)	Different	Yes	[41]

[a]Method: VP, variance partitioning approach; TB, trait-based approach.

[b]Phenotype of social interaction partner affecting the phenotype of a focal individual: (a) the among-individual, and (w) the within-individual component of the phenotype of the social partner can result in (a) among-individual, and (w) within-individual variation in the phenotype of a focal individual; →, non-reciprocal effect; ↔, reciprocal effect; 'unknown' is printed when the phenotype of the social partner (either representing effects of a single trait or of combinations of traits) was not quantified.

[c]Observed variation in the phenotype of the focal was due to within-partner variation in phenotype across repeated interactions with the same partner ('Same') versus among-partner variation in phenotype across interactions with different partners ('Different').

[d]Variation in Ψ: presence of variation in social responsiveness (Ψ) considered at any level.

Interactions with multiple social partners

Individuals commonly interact with multiple conspecifics, for example, in competitive contexts. Quantitative geneticists have developed pointed approaches to studying IGEs beyond simply dyadic interactions (which largely go beyond the scope of this paper). In short, the trait-based approach (Box 3) might, for example, use an aggregate statistic of a group of individuals (cf. mean or variance) as its predictor variable [43]. One might also model situations where each individual is exposed to different groups or different group compilations within a reaction-norm framework. Similarly, the variance-partitioning approach (Box 2) can be extended to include interactions between groups of more than two individuals of equal [44] or variable group size [45,46]. Alternatively, one can weigh by number (or intensity) of interactions between sets of interacting individuals, an approach used in forestry science, where the intensity of competition between individual trees is modelled by incorporating a distance matrix in the statistical analysis [47]. For animal studies, interaction metrics (e.g., derived from social network analysis [48]) can similarly be included to incorporate interactions between multiple individuals.

Types of social environment effects

Social environment effects exist whenever phenotypes of interaction partners affect the phenotype of a focal individual (Figure 1). Such socially responsive phenotypes are sometimes termed 'interactive phenotypes' in quantitative genetics [17]. They exist in a variety of forms, ranging from behaviours that are marginally influenced by phenotypes of partners to those that cannot be defined outside the context of social interactions. Variation in lay-date in birds is, for example, largely attributable to the characteristics of the female, year, and breeding location, rather than to those of her male partner [49]. By contrast, aggressiveness or cooperative tendencies are only expressed as part of interactions. Social environment effects can be classified as reciprocal versus nonreciprocal [17]. Nonreciprocal interactions equate to RNs where each phenotypic trait is uniquely defined as either predictor or response; for example, the aggressiveness of an individual is a function of the size of an opponent but the size of the individual is not necessarily a function of the level of aggressiveness of the opponent. By contrast, reciprocal interactions equate to RNs where the predictor phenotype of one individual represents the response variable of another individual (Table 1, [50]). The aggressiveness of an individual might, for instance, elicit aggressiveness in opponents [26], causing feedbacks between phenotype and social environment [17]. Such feedback loops can either increase or decrease the amount of among-individual variation [51]. Some interactive phenotypes, finally, represent characteristics of interactions rather than individuals, such as duration or intensity of interactions, or other emergent characters [43] such as the productivity or aggressiveness of ant colonies (e.g., [52]). Understanding how the different types of interactions among individuals affect phenotypic variation is key to both behavioural ecology and quantitative genetics.

Box 2. Focal and partner identity effects due to repeatable phenotypes

Variation in phenotype of a focal individual due to among-partner variation in phenotype can be estimated with a variance-partitioning approach [69]. This requires datasets (i) where focals are repeatedly assayed for their behaviour, (ii) where focals switch partners across assays, and (iii) where each partner occurs (to some extent) with multiple focals (Figure 2B). This could apply to female birds that switch mates across breeding attempts [49] or to male rodents that switch opponents across aggressive interactions [26]. Cross-classified mixed-effect models can be applied to such data, and random intercepts for focal and partner identity can be fitted to decompose the overall phenotypic variance (V_P) into variance due to the identity of the focal (V_{I_D}; I for individual D for direct effect), the identity of its social partner (V_{I_S}; S for social), versus unexplained variance (V_R; R for residual), where $V_P = V_{I_D} + V_{I_S} + V_R$. Here, the familiar individual repeatability (R) equals $\frac{V_{I_D}}{V_P}$ [31]; the proportion of variance caused by repeatable aspects of the phenotype of a partner equals $\frac{V_{I_S}}{V_P}$ [26]. In this approach the phenotype of the partner that is affecting the behaviour of a focal individual remains unidentified, but variance attributable to partner identity ($V_{I_S} > 0$) nevertheless implies that its phenotype does matter (Box 3).

Partitioning variation in this way provides considerable insight. Imagine that we had collected an observational dataset of the same type but we had not controlled for who interacted with whom, a common situation in field studies. Suppose further that we had failed to consider social environment effects (V_{I_S}) and concluded, based on a model with only random intercepts for the identity of the focal individual (i.e., $V_P = V_{I_D} + V_R$), that the behaviour of the focal showed substantial among-individual variance. Given the observational nature of the data, another conclusion would have been possible: individuals appeared more different than they were because their social partner choice was repeatable, causing apparent among-individual variation [31,36]. The extent to which this was the case could be investigated by including random intercepts for partner identity into the mixed-effects model. If individual repeatability was partly due to social environment repeatability, partner identity (V_S) should explain variance previously attributed to the identity of the focal (V_{I_D}), and individual repeatability should consequently decrease. Another question that would emerge from such exercises is what proportion of the within-focal variation (cf. $V_{I_S} + V_R$) was explained by within-focal variation in social environment (V_{I_S}). If present, it implies that individuals were partly plastic with respect to social context.

Integrating quantitative genetics and behavioural ecology theory

Quantitative geneticists have developed various techniques for quantifying social influences on phenotypes [53]. Their research utilises individuals that vary in relatedness (cf. 'pedigreed' populations [54]) owing to their interest in additive genetic (co)variances [22]. In its simplest form, focal individuals are phenotyped once with a social interaction partner (Figure 2A), followed by the application of one of two main approaches [53]. In the 'trait-based' approach, the presumed phenotype (P_k) of the partner k affecting the phenotype (P_j) of a focal individual j is assayed, and mixed-effect 'animal' [54] models are fitted to estimate the additive genetic (A_j) and residual (e_j) merits of each individual, while quantifying the population-level dependency (Ψ) of the phenotype of the focal on that of its partner ($cf.\ P_j = A_j + \psi P_k + e_j$) [17]. For nonreciprocal interactions, the 'interaction coefficient' Ψ simply represents an ordinary least-square regression coefficient [53]. For reciprocal interactions, calculation of Ψ is more complex (in terms of parameters to be estimated) because phenotypes of interacting individuals feedback on each other ([53] for modelling solutions). In the alternative 'variance-partitioning' approach (Box 2), the phenotype of the partner remains unquantified, but variances attributable to DGEs (V_{A_D}), IGEs (V_{A_S}) and their correlation (r_{A_D,A_S}) are estimated. The logic is that if phenotypes of partners affect the behaviour of a focal individual (cf. $\Psi \neq 0$), IGEs explain variation (i.e., $V_{A_S} > 0$). The two approaches are complementary (Box 3): parameters of variance-component approaches (V_{A_D}, V_{A_S} and r_{A_D,A_S}) can be calculated from those estimated with trait-based models (V_{A_j} and Ψ) [53,55], but typically not vice versa.

Box 3. Individual responsiveness to among- versus within-partner variation in phenotype

Variance-partitioning approaches (Box 2) estimate variance attributable to partner identity [55]. This source of variation is caused by among-individual variance in partner phenotype. Presumed effects of identified candidate partner phenotypes can be estimated with a trait-based approach [53], which we apply here to the example in Box 2. Whereas we previously partitioned phenotypic variance in the behaviour of the focal (V_P) into among-focal (V_{I_D}), among-social partner (V_{I_S}), and residual variance (V_R) (Box 2), here we evaluate whether identified among-individual characteristics of partners (e.g., their size) can explain the social environment effect (V_{I_S}). Such hypotheses are tested by fitting partner phenotype as a fixed-effect covariate into the model (estimating Ψ [53]). If the trait was fully responsible for the partner identity effect (V_{I_S}), its inclusion should reduce this component to zero.

In this opinion article, we argue that trait-based approaches should distinguish between responses to predictable versus plastic components of partner phenotypes because theory often addresses one of the two specifically [13,15]. Imagine that a labile phenotype (P_{ik}) of partners (k) was assayed during each observation (i) of the focals' behaviour (P_{ij}) (as in Figure 2B, main text). Instead of estimating Ψ by fitting the 'raw' phenotype of the partner (P_{ik}) as a covariate [53], one might fit the mean phenotype of the partner over all observations ($\bar{P}_k$), and deviation from this mean during a focal instance (i.e., $P_{ik} - \bar{P}_k$), as covariates into the mixed-effects model (with random intercepts for the identities of the focal and the social partner). This enables estimation of phenotypic plasticity in focals as a function of repeatable ($\psi_A \bar{P}_k$; A for among-partner) and plastic ($\psi_W(P_{ik} - \bar{P}_k)$; W for within-partner) components of partner phenotypes (Equation I):

$$P_{ij} = \mu + I_{D0j} + I_{S0k} + \psi_A \bar{P}_k + \psi_W (P_{ik} - \bar{P}_k) + e_{0ij} \quad [I]$$

Here, the focal phenotype (P_{ij}) is decomposed into the RN elevation of the focal (I_{D0j}) and contribution of partner identity (I_{S0k}) (both expressed as deviations from the population-average elevation, μ), two RN slopes [the dependence of its phenotype on repeatable (Ψ_A) and plastic (Ψ_W) parts of the phenotype of its partner], and the deviation of each instance from its estimated RN (e_{0ij}). This model, notably, only returns population-average estimates of social responsiveness (Ψ_A, Ψ_W), and can be expanded to estimate among-individual variation in social responsiveness by including random slopes [42] for the identity of the focal (Equation II):

$$P_{ij} + \mu + I_{D0j} + I_{S0k} + (\psi_A + I_{1jA})\bar{P}_k + (\psi_W + I_{1jW})(P_{ik} - \bar{P}_k) + e_{0ij} \quad [II]$$

Doing so enables empirical testing of behavioural ecology theory [13–15] that predicts among-individual differences in responsiveness to predictable components of the phenotype of the partner ($+I_{1jA}$) versus among-individual differences in responsiveness to moment-to-moment changes in partner phenotype ($+I_{1jW}$).

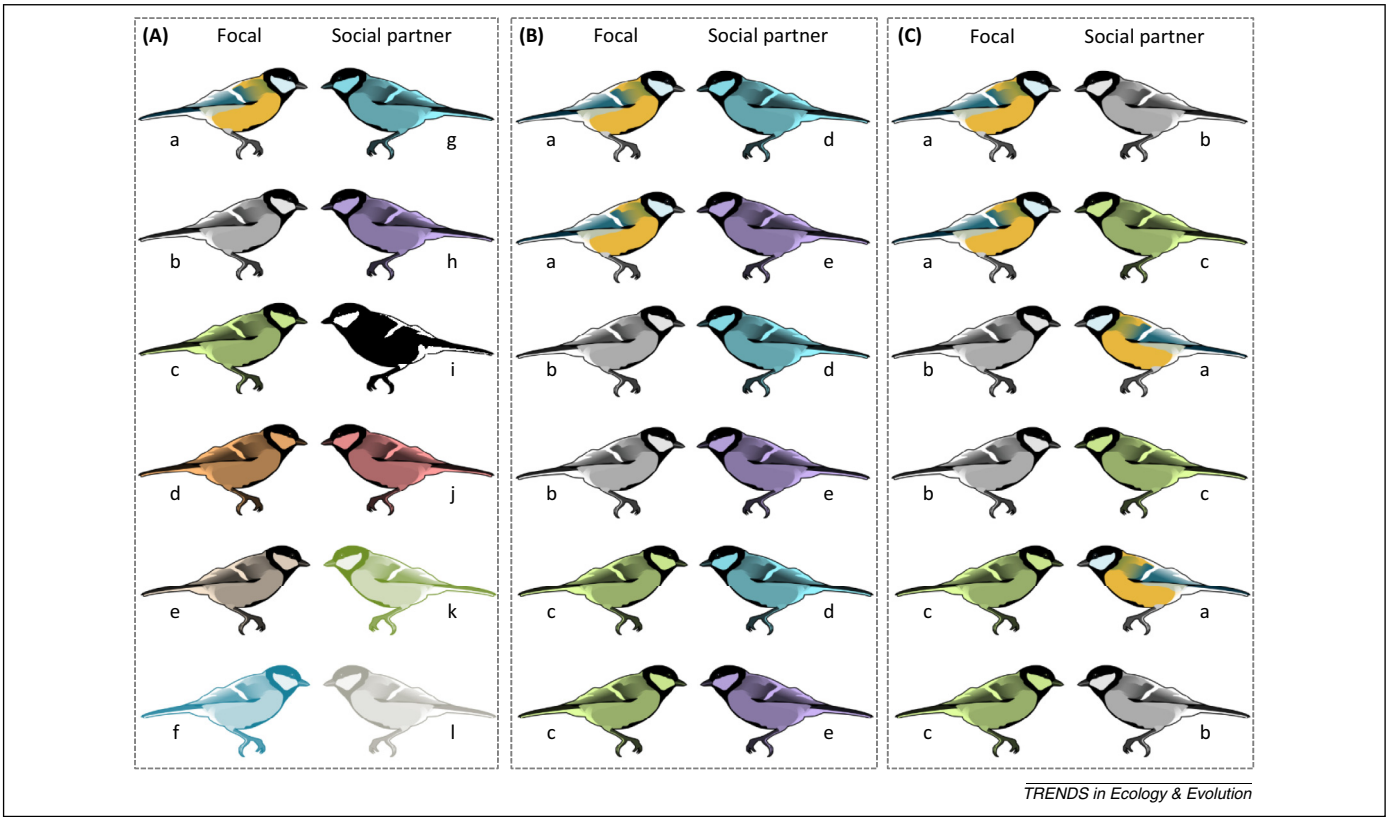

Figure 2. Study designs enabling the estimation of focal and partner identity effects. **(A)** Particular individuals (lower-case letters) are assigned 'focal' (a–f), others 'social partner' (g–l); all individuals are used once. This classic quantitative genetic design combined with a pedigreed population enables the estimation of variance among focals attributable to direct genetic effects (V_{A_D}), indirect genetic effects (V_{A_S}), and their genetic correlation (r_{A_D,A_S}); the level of social responsiveness (cf. interaction coefficient Ψ) can be estimated provided that the phenotype of the partner was also assayed. **(B)** Particular individuals (lower-case letters) are assigned 'focal' (a–c), others 'social partner' (d–e); for example, lay-dates are measured for focal females and their male partners [23,49]. Each focal individual is assayed repeatedly, but always with another social partner, and each social partner occurs with multiple focals. This design enables estimation of variance attributable to focal (V_{I_D}) and partner identity (V_{I_S}) effects (Box 2); those effects can be further partitioned to estimate V_{A_D}, V_{A_S}, and r_{A_D,A_S} provided that pedigree information is available. The level of social responsiveness (cf. interaction coefficient Ψ) can be estimated provided that the phenotype of the partner had been assayed (Box 3). Study designs where both individuals were assayed in each round, furthermore, enable testing social niche hypotheses as detailed in Box 4. **(C)** Similar to scenario (B) but applicable to situations where only the behaviour of the focal individual is assayed but where each individual plays either focal or social partner depending on the round; for example, aggressive behaviour is assayed in each round but only for the individual assigned focal [26]. Variance components detailed under (A) and (B) are estimable, as is the individual-level correlation (r_{I_D,I_S}) between the focal and social partner identity effects; r_{I_D,I_S} would, for example, be positive when aggressive individuals also elicit aggressiveness in others [26].

The quantitative genetic equation specifying the dependency of the phenotype of an individual j on a phenotype of a partner k (cf. $P_j = A_j + \psi P_k + e_j$) greatly resembles the way in which behavioural ecologists model the dependency of a behavioural phenotype on the environment across observations i of a single individual (cf. its RN: $P_{ij} = I_j + \beta_j x_{ij} + e_{ij}$). Expanding RNs to include the phenotype of a conspecific as an environmental axis that varies within a single individual, as proposed here, merges the two approaches (cf. $P_{ij} = I_j + \psi_j P_{ik} + e_{ij}$); here, the quantitative genetic interaction coefficient Ψ corresponds to the slope of the RN of an individual (cf. level of plasticity or 'social responsiveness'). Whereas the existence of heritable variation in social responsiveness (Ψ) represents an area of growing interest in quantitative genetics [28,29], adaptive theory in behavioural ecology is currently revealing the ecological conditions that favour versus disfavour the evolution of both social responsiveness and individual differences in social responsiveness [11]. One fascinating notion inherent to viewing indirect effects in terms of RNs is the possibility of genetic correlations between elevations and slopes, as suggested in the stress-physiology literature [40]. This is illustrated in Figure 1B where the phenotypic/genetic variance among focal individuals, as well as repeatability/heritability, decreases with increasing value of the phenotype of the partner because of a negative elevation–slope correlation. Such a genetic architecture would logically result in a changing importance of IGEs as a source of phenotypic variance across the social gradient. Quantitative genetic theory has yet to address the evolutionary repercussions of such genetic characteristics of individual RNs, underlining the great potential for cross-fertilisation between these distinct fields of biology.

Multilevel variation in responsiveness

Recent adaptive theory implies that social interactions can favour the evolution of (i) within-individual responsiveness in behaviour to predictable components of behavioural phenotypes of social partners (cf. their personality), and (ii) among-individual differentiation in level of responsiveness [11]. Moreover, individuals are also generally predicted to respond behaviourally to changes in behaviour within their social partners [12], for example when negotiating how much each party might invest in parental care. Therefore, owing to the characteristic multilevel nature of behaviour [1,2], predictions regarding social responsiveness require the partitioning of partner effects into those caused by repeatable (cf. predictable) versus changeable

(cf. plastic) components of partner phenotypes. This partitioning might be achieved by applying study designs that explicitly acknowledge the existence of multilevel variation in partner phenotype, followed either by a variance partitioning (Box 2) or trait-based analytical approach (Box 3). The variance partitioning approach might, for example, be applied to test whether any predictable (cf. repeatable), although unidentified, component of a phenotype of a partner affected the behaviour of the focal. This requires a study design where each social partner occurs with multiple focal individuals (Figure 2B,C) such that the existence of a partner identity effect can be estimated (Box 2). In cases where each assay included the quantification of the phenotype of the partner that was hypothesised to affect the behaviour of the focal, centring techniques [31] can subsequently be applied to partition population-average estimates of responsiveness (Ψ) into responsiveness to the repeatable (Ψ_A; A for among) and plastic (Ψ_W; W for within) part of the phenotype of the partner (Box 3). Finally, adaptive theory regarding individual differences

Box 4. Testing predictions of the social niche hypothesis

The social niche hypothesis proposes alternative strategies favoured by negative frequency-dependent selection in how animals trade-off investment in costly behaviours [9]. Parents, for instance, must invest in both nest defence and offspring provisioning [57], and selection thus favours division of labour. Specific predictions of the hypothesis are phrased here in statistical terms, applying a hybrid [56] between variance partitioning (Box 2) and trait-based approaches (Box 3). For simplicity, we assume that individuals do not differ in responsiveness.

Imagine that male and female partners were assayed for antipredator aggressiveness during each breeding attempt, and that each parent bred multiple times although not always with the same partner (as in Figure 2B, main text). Three distinct predictions can now be formulated concerning how social niche specialisation might emerge. Prediction 1: within pairs across breeding attempts, within-male upregulation in aggressiveness associates with within-female downregulation in aggressiveness (Figure IA). Prediction 2: females are repeatable and males respond plastically to the repeatable part of the behaviour of the female; males mated with aggressive female types downregulate their aggressiveness but those same males upregulate their aggressiveness when switching to breed with a less-aggressive type (Figure IB) (the same scenario applies to the opposite sex). Prediction 3: both sexes show among-individual variation in aggressiveness, and types are disassortatively paired. The first two predictions represent parameters of a bivariate mixed-effects model with female and male behaviour as the two response variables and random intercepts for female and male identity [31,54,56]. Phenotypic variance in female behaviour (V_{P_F}) is thus partitioned into a female identity (V_F), male identity (V_M), and a residual (V_R) component (i.e., $V_P = V_F + V_M + V_R$); the same applies to male behaviour (V_{P_M}). Covariances between male and female behaviour are also estimated at the female identity (Cov_F), male identity (Cov_M), and residual levels (Cov_R). Prediction 1 implies a negative within-pair residual covariance ($Cov_R < 0$): in a year where a male upregulates its aggressiveness its mate downregulates hers (Figure IA). Prediction 2 implies a female identity effect ($V_F > 0$) in female behaviour and a negative covariance between male and female behaviour at the female identity level ($Cov_F < 0$; cf. a plastic response of the male to the repeatable part of the behaviour of the female): females that are on average relatively aggressive reduce aggressiveness in their mates. A male identity effect ($V_M > 0$) in male behaviour, and a negative covariance between male and female behaviour at the male identity level ($Cov_M < 0$), are similarly expected.

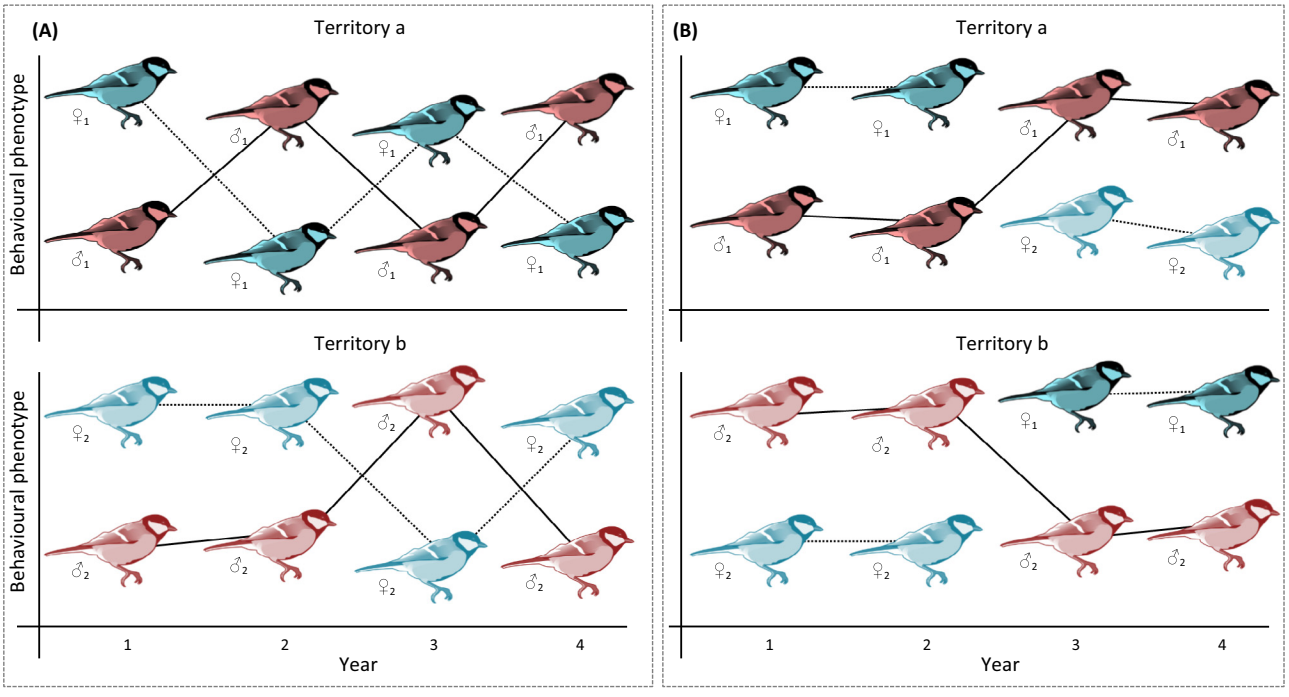

Figure I. Predictions of the social niche hypothesis illustrated graphically. **(A)** ♂1 and ♂2 breed for four consecutive years on territories a and b, respectively. ♂1 is always paired with ♀1; ♂2 is always paired with ♀2. Division of labour is achieved because males upregulate their behaviour in years where their mate downregulates hers, causing a negative within-pair residual covariance ($Cov_R < 0$). **(B)** Modified scenario where ♂1 and ♂2 switch mates after year 2. Here, females are repeatable (cf. ♀1 is relatively aggressive in all four years compared with ♀2) and males are plastically responding to female type: both males are relatively unaggressive when mated with ♀1 but upregulate their aggressiveness when mated with ♀2. This causes a negative covariance between male and female behaviour at the female identity level ($Cov_F < 0$).

in responsiveness can be tested empirically by applying statistical techniques that enable the estimation of among-individual variance in social responsiveness in either type (Ψ_A, Ψ_W; Box 3). The existence of among-individual variance in responsiveness would also provide clues about the potential for heritable variation in responsiveness in the population, a key question in quantitative genetics [34,54].

Social niche hypotheses

Acknowledgement of multilevel variation in social responsiveness is, in our opinion, also key in testing predictions of seemingly unrelated hypotheses, such as those proposing that social interactions lead to social niche specialisation [9,10]. In Box 2, we detail how a hybrid [56] between the variance partitioning and trait-based approaches might be applied to address empirically social niche hypotheses. The proposed approach enables testing of a specific set of predictions that fully acknowledge the multilevel nature of phenotypes of individuals and their social environments (Box 4). Specifically, we would expect that the repeatable part of the phenotype of a social partner is negatively associated within the plastic part of the same phenotype of a focal individual (see Figure IB in Box 4). For example, a male paired with an aggressive personality type might trade-off time invested in territory defence in favour of parental care [57,58]. The same male might upregulate investment in territory defence after switching to breed with a less-aggressive female personality type. In other words, the among-individual component of the behaviour of the partner causes within-individual adjustment in the focal individual, enabling division of labour within the two cooperating individuals. Another way in which such a division might be achieved is by means of simultaneous adjustments in behaviour across the two partners (see Figure IA in Box 4). For example, males that upregulate investment in territorial defence from one breeding attempt to the next might elicit their mates to downregulate investment in territorial defence, leading to a negative covariance between the plastic part of the same behaviour across two partners. Testing such predictions requires study designs where the same combination of focal individual and social partner is repeatedly assayed for their phenotype, while partners simultaneously occur with multiple focal individuals (Figure 2B). In other words, the fascinating biology caused by social interactions might best be revealed by fully acknowledging the multilevel nature of behaviour when addressing biological hypotheses.

Concluding remarks

Animal behaviour varies across multiple levels [1]. Individuals differ in average behaviour, show phenotypic plasticity, and vary in level of plasticity [2]. In this paper we have highlighted recent theory in behavioural ecology predicting the existence of individual variation in 'social responsiveness' to among- and within-individual components of conspecific phenotypes [9,11,13]. We have also summarised how social responsiveness has been widely addressed in the field of quantitative genetics through the study of indirect genetic effects [17,18]. We further detailed how behavioural ecologists might benefit from adopting approaches developed in quantitative genetics to test adaptive theories (Boxes 2–4), while quantitative genetics might benefit from behavioural ecology theory regarding the ecological conditions favouring (variation in) social responsiveness (Box 1), and distinct types of responsiveness that might exist in nature (Box 3). Owing to adaptive theory predicting responsiveness to predictable versus plastic components of partner phenotype [9,11,13], we propose that research should appreciate and quantify effects of social environments that vary across multiple

Box 5. Outstanding questions

Evolution of responsiveness

What are the ecological conditions that favour responsiveness towards repeatable (Ψ_A) versus plastic (Ψ_W) parts of partner phenotypes, and in what types of behaviour? Which ecological conditions generate the evolution of responsiveness of each type rather than the evolution of general responsiveness (Ψ)?

RN structure and variation

How common are (genetic) correlations between average phenotype (RN elevation; 'personality') and level of social responsiveness (RN slope; plasticity; Ψ)? How do genetic correlations between these two RN components affect evolutionary trajectories? In what types of behaviour do individuals or genotypes vary in social responsiveness (Ψ), and how is this variation maintained?

Emergent properties

Emergent properties, such as the intensity of a fight (Figure IA) or the productivity of a colony (Figure IB), are a function of characteristics of interacting individuals, such as their respective size [37] (Figure IA) or aggressiveness [52] (Figure IB). When applying a RN framework to interactions rather than individuals, how does the very presence of among- versus within-individual variation in interacting individuals affect such emergent properties [70]?

Social networks

Individuals differ tremendously in social behaviour, reflected in interaction intensity and number of interaction partners [71]. Social network characteristics (e.g., 'degree' or 'betweenness') vary among individuals as a function of their personality [72], but how does social responsiveness affect social network structure? Does selection favour the integration of individual network characteristics and level of social responsiveness?

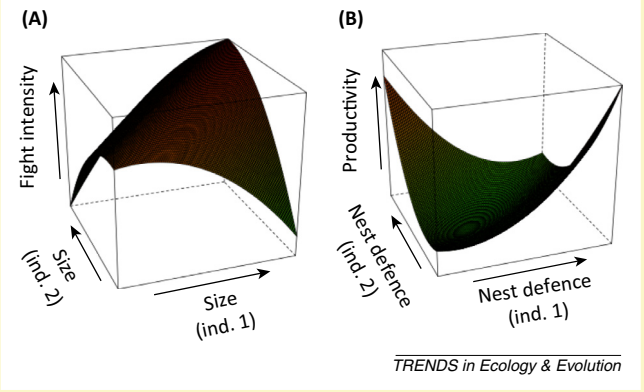

Figure I. Emergent properties as a function of individual-level characteristics. Fight intensity might be increased **(A)** but colony productivity decreased **(B)** as a function of the level of phenotypic similarity between interacting individuals (ind.).

levels. Doing so would help further integrate distinct areas of evolutionary biology [30,32] and address outstanding questions (Box 5) regarding the evolution and ecology of labile phenotypic characters (cf. behaviour) that vary explicitly among and within individuals.

Acknowledgements

We gratefully acknowledge input by Susanne Foitzik, Chang Seok Han, Petri Niemelä, Francesca Santostefano, Max Wolf, Alastair Wilson, two anonymous reviewers, and participants of the Symposium 'Personality: Causes and Consequences of Consistent Behavioural Variation' funded by the Volkswagen Foundation. Figure 2 and Figure I in Box 4 featuring great tit *Parus major* individuals are modified from the public domain website http://clipartist.net/.

References

1 Araya-Ajoy, Y.G. and Dingemanse, N.J. (2014) Characterizing behavioural 'characters': a conceptual and statistical framework. *Proc. R. Soc. Lond. B: Biol. Sci.* 281, 20132645
2 Dingemanse, N.J. *et al.* (2010) Behavioural reaction norms: animal personality meets individual plasticity. *Trends Ecol. Evol.* 25, 81–89
3 Houston, A.I. and McNamara, J.M. (1999) *Models of Adaptive Behaviour*, Cambridge University Press
4 Dingemanse, N.J. and Wolf, M. (2010) A review of recent models for adaptive personality differences. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 365, 3947–3958
5 Wolf, M. and Weissing, F.J. (2010) An explanatory framework for adaptive personality differences. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 365, 3959–3968
6 Dingemanse, N.J. and Wolf, M. (2013) Between-individual differences in behavioural plasticity: causes and consequences. *Anim. Behav.* 85, 1031–1039
7 Wolf, M. and McNamara, J.M. (2012) On the evolution of personalities via frequency-dependent selection. *Am. Nat.* 179, 679–692
8 Sih, A. and Del Giudice, M. (2012) Linking behavioural syndromes and cognition: a behavioural ecology perspective. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 367, 2762–2772
9 Bergmüller, R. and Taborsky, M. (2010) Animal personality and social niche specialisation. *Trends Ecol. Evol.* 25, 504–511
10 Montiglio, P.O. *et al.* (2013) Social niche specialization under constraints: personality, social interactions and environmental heterogeneity. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 368, 20120343
11 Wolf, M. and McNamara, J.M. (2013) Adaptive between-individual differences in social competence. *Trends Ecol. Evol.* 28, 253–254
12 Maynard Smith, J. (1982) *Evolution and the Theory of Games*, Cambridge University Press
13 Dall, S.R.X. *et al.* (2004) The behavioural ecology of personality: consistent individual differences from an adaptive perspective. *Ecol. Lett.* 7, 734–739
14 McNamara, J.M. *et al.* (2009) Evolution of trust and trustworthiness: social awareness favours personality differences. *Proc. R. Soc. Lond. B: Biol. Sci.* 276, 605–613
15 Wolf, M. *et al.* (2011) On the coevolution of social responsiveness and behavioural consistency. *Proc. R. Soc. Lond. B: Biol. Sci.* 278, 440–448
16 Johnstone, R.A. and Manica, A. (2011) Evolution of personality differences in leadership. *Proc. Natl. Acad. Sci. U.S.A.* 108, 8373–8378
17 Moore, A.J. *et al.* (1997) Interacting phenotypes and the evolutionary process. 1. Direct and indirect genetic effects of social interactions. *Evolution* 51, 1352–1362
18 Wolf, J.B. *et al.* (1999) Interacting phenotypes and the evolutionary process. II. Selection resulting from social interactions. *Am. Nat.* 153, 254–266
19 Bijma, P. and Wade, M. (2008) The joint effects of kin, multilevel selection and indirect genetic effects on response to genetic selection. *J. Evol. Biol.* 21, 1175–1188
20 McGlothlin, J.W. *et al.* (2010) Interacting phenotypes and the evolutionary process. III. Social evolution. *Evolution* 64, 2558–2574
21 Wolf, J.B. *et al.* (1998) Evolutionary consequences of indirect genetic effects. *Trends Ecol. Evol.* 13, 64–69
22 Lynch, M. and Walsh, B. (1998) *Genetics and Analysis of Quantitative Traits*, Sinauer
23 Brommer, J.E. and Rattiste, K. (2008) 'Hidden' reproductive conflict between mates in a wild bird population. *Evolution* 62, 2326–2333
24 Westneat, D.F. (2011) Evolution in response to social selection: the importance of interactive effects of traits on fitness. *Evolution* 66, 890–895
25 Wolf, J.B. (2003) Genetic architecture and evolutionary constraint when the environment contains genes. *Proc. Natl. Acad. Sci. U.S.A.* 100, 4655–4660
26 Wilson, A.J. *et al.* (2009) Indirect genetic effects and the evolution of aggression in a vertebrate system. *Proc. Biol. Sci.* 276, 533–541
27 Bailey, N.W. and Moore, A.J. (2012) Runaway sexual selection without genetic correlations: social environments and flexible mate choice initiate and enhance the Fisher process. *Evolution* 66, 2674–2684
28 Bailey, N.W. and Zuk, M. (2012) Socially flexible female choice differs among populations of the Pacific field cricket: geographical variation in the interaction coefficient psi (ψ). *Proc. R. Soc. Lond. B: Biol. Sci.* 279, 3589–3596
29 Chenoweth, S.F. *et al.* (2010) Experimental evidence for the evolution of indirect genetic effects: changes in the interaction effect coefficient, psi (ψ), due to sexual selection. *Evolution* 64, 1849–1856
30 Smiseth, P.T. *et al.* (2008) Parent–offspring conflict and co-adaptation: behavioural ecology meets quantitative genetics. *Proc. R. Soc. Lond. B: Biol. Sci.* 275, 1823–1830
31 Dingemanse, N.J. and Dochtermann, N.A. (2013) Quantifying individual variation in behaviour: mixed-effect modelling approaches. *J. Anim. Ecol.* 82, 39–54
32 Dingemanse, N.J. and Dochtermann, N.A. (2014) Individual behaviour: behavioural ecology meets quantitative genetics. In *Quantitative Genetics in the Wild* (Charmantier, A. *et al.*, eds), pp. 54–67, Oxford University Press
33 Grafen, A. (1984) Natural selection, kin selection and group selection. In *Behavioural Ecology: An Evolutionary Approach* (Krebs, J.R. and Davies, N.B., eds), pp. 62–84, Blackwell Scientific
34 Nussey, D.H. *et al.* (2007) The evolutionary ecology of individual phenotypic plasticity in wild populations. *J. Evol. Biol.* 20, 831–844
35 Mathot, K. *et al.* (2011) Disentangling the roles of frequency- versus state-dependence in generating individual differences in behavioural plasticity. *Ecol. Lett.* 14, 1254–1262
36 Westneat, D.F. *et al.* (2011) Individual variation in parental care reaction norms: integration of personality and plasticity. *Am. Nat.* 178, 652–667
37 Wilson, A.J. *et al.* (2011) Integrating animal personality research and animal context theory: aggressiveness in the green swordtail *Xiphophorus helleri*. *Plos ONE* 6, 1–13
38 Meaney, M.J. (2001) Maternal care, gene expression, and the transmission of individual differences in stress reactivity across generations. *Annu. Rev. Neurosci.* 24, 1161–1192
39 Mathot, K.J. *et al.* (2012) Adaptive strategies for managing uncertainty may explain personality-related differences in behavioural plasticity. *Oikos* 121, 1009–1020
40 Coppens, C.M. *et al.* (2010) Coping styles and behavioural flexibility: towards underlying mechanisms. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 365, 4021–4028
41 Natarajan, D. *et al.* (2009) Delineation of violence from functional aggression in mice: an ethological approach. *Behav. Gen.* 39, 73–90
42 Schaeffer, L.R. (2004) Application of random regression models in animal breeding. *Livest. Prod. Sci.* 86, 35–45
43 Lloyd, E.A. and Gould, S.J. (1993) Species selection on variability. *Proc. Natl. Acad. Sci. U.S.A.* 90, 595–599
44 Bijma, P. *et al.* (2007) Multilevel selection 1: quantitative genetics of inheritance and response to selection. *Genetics* 175, 277–288
45 Hadfield, J.D. and Wilson, A.J. (2007) Multilevel selection 3: modeling the effects of interacting individuals as a function of group size. *Genetics* 177, 667–668
46 Bijma, P. (2010) Multilevel selection 4: modeling the relationship of indirect genetic effects and group size. *Genetics* 186, 1029–1031
47 Costa e Silva, J. and Kerr, R.J. (2013) Accounting for competition in genetic analysis, with particular emphasis on forest genetic trials. *Tree Genet. Genomes* 9, 1–17
48 Wey, T. *et al.* (2008) Social network analysis of animal behaviour: a promising tool for the study of sociality. *Anim. Behav.* 75, 333–344

49 Browne, W.J. *et al.* (2007) Using cross-classified multivariate mixed response models with application to life history traits in great tits (*Parus major*). *Stat. Model.* 7, 217–238
50 Trubenová, B. and Hager, R. (2012) Phenotypic and evolutionary consequences of social behaviours: interactions among individuals affect direct genetic effects. *Plos ONE* 7, e46273
51 Sih, A. *et al.* (2015) Animal personality and state-behaviour feedbacks: a review and guide for empiricists. *Trends Ecol. Evol.* 30, 50–60
52 Scharf, I. *et al.* (2012) Characterizing the collective personality of ant societies: aggressive colonies do not abandon their home. *PLoS ONE* 7, e33314
53 Bijma, P. (2014) The quantitative genetics of indirect genetic effects: a selective review of modelling issues. *Heredity* 112, 61–69
54 Wilson, A.J. *et al.* (2010) An ecologist's guide to the animal model. *J. Anim. Ecol.* 79, 13–26
55 McGlothlin, J.W. and Brodie, E.D. (2009) How to measure indirect genetic effects: the congruence of trait-based and variance-partitioning approaches. *Evolution* 63, 1785–1795
56 McAdam, A.G. *et al.* (2014) The effects of others' genes: maternal and other indirect genetic effects. In *Quantitative Genetics in the Wild* (Charmantier, A. *et al.*, eds), pp. 84–103, Oxford University Press
57 Mutzel, A. *et al.* (2013) Temporal trade-offs between nestling provisioning and defence against nest predators in blue tits. *Anim. Behav.* 85, 1459–1469
58 Mutzel, A. *et al.* (2013) Parental provisioning behaviour plays a key role in linking personality with reproductive success. *Proc. R. Soc. Lond. B: Biol. Sci.* 280, 20131019
59 Houston, A.I. (2013) Evolutionary games and gambits. *Behav. Ecol.* 24, 12
60 Westneat, D.F. *et al.* (2014) The biology hidden inside residual within-individual phenotypic variation. *Biol. Rev.* Published online July 30, 2014. http://dx.doi.org/10.1111/brv.12131
61 Luttbeg, B. and Sih, A. (2010) Risk, resources, and state-dependent adaptive behavioural syndromes. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 365, 3977–3990
62 Nicolaus, M. *et al.* (2012) Experimental evidence for adaptive personalities in a wild passerine bird. *Proc. R. Soc. Lond. B: Biol. Sci.* 279, 4885–4892
63 Laskowski, K.L. and Pruitt, J.N. (2014) Evidence of social niche construction: persistent and repeated social interactions generate stronger personalities in a social spider. *Proc. R. Soc. Lond. B: Biol. Sci.* 281, 20133166
64 Moshgani, M. and Van Dooren, T.J. (2011) Maternal and paternal contributions to egg size and egg number variation in the blackfin pearl killifish *Austrolebias nigripinnis*. *Evol. Ecol.* 25, 1179–1195
65 Brooks, R. and Endler, J.A. (2001) Female guppies agree to differ: phenotypic and genetic variation in mate-choice behavior and the consequences for sexual selection. *Evolution* 55, 1644–1655
66 Daisley, J.N. *et al.* (2005) Enhanced yolk testosterone influences behavioral phenotype independent of sex in Japanese quail chicks *Coturnix japonica*. *Horm. Behav.* 47, 185–194
67 Kölliker, M. *et al.* (2000) The quantitative genetic basis of offspring solicitation and parental response in a passerine bird with parental care. *Proc. R. Soc. Lond. B: Biol. Sci.* 267, 2127–2132
68 van Oers, K. *et al.* (2008) Personality is associated with extrapair paternity in great tits. *Parus Major. Anim. Behav.* 76, 555–563
69 Griffing, B. (1967) Selection in reference to biological groups. I. Individual and group selection applied to populations of unordered groups. *Aust. J. Biol. Sci.* 20, 127–142
70 Modlmeier, A.P. and Foitzik, S. (2011) Productivity increases with variation in aggression among group members in *Temnothorax* ants. *Behav. Ecol.* 22, 1026–1032
71 Kurvers, R.H.J.M. *et al.* (2014) The evolutionary and ecological consequences of animal social networks: emerging issues. *Trends Ecol. Evol.* 29, 326–335
72 Wilson, A.D.M. *et al.* (2013) Network position: a key component in the characterization of social personality types. *Behav. Ecol. Sociobiol.* 67, 163–173

Chapter 4

Evolutionary constraints in the alternative routes to fertilization success of male Great tits: trade-offs and indirect female effects.

Yimen G. Araya-Ajoy, Bart Kempenaers, Sylvia Kuhn, Kimberly J. Mathot, Alexia Mouchet, Ariane Mutzel, Marion Nicolaus, Jan Wijmenga, Niels J. Dingemanse

Manuscript

Abstract

Extra-pair reproduction in socially monogamous species creates alternative siring routes for males. They can increase their siring success via three routes: mating with highly fecund females, avoiding within-pair paternity loss or seeking for extra-pair fertilizations. In natural populations, variation in traits associated to these routes is pervasive, which is puzzling, because selection should deplete variation in traits closely linked to fitness. In this study we focus on two aspects that could explain the maintenance of variation in male siring routes; social environment effects (through social female effects) and the existence of trade-offs between routes. In wild populations, the quantification of such processes is challenging because trade-offs can exist at various hierarchical levels (among-individuals *and* within-individuals), while each route is difficult to define as the property of just one member of the social pair. We studied this complexity by quantifying variation in male siring success of wild great tits across four consecutive years. We used a variance partitioning approach and studied male siring success and its routes as “interactive phenotypes” arising from phenotypic contributions of both members of the social pair. We found that the female mate has a substantial effect on male siring success, mostly through her fecundity but also through her promiscuity. We then proceeded to study the relationships between the different siring routes to determine whether covariation within- or among-individual males were consistent with the existence of a trade-off. We show that nests with larger clutches had more chicks sired by an extra pair male, causing a trade-off between a male’s within-pair paternity loss and his social female fecundity. Males that consistently gained more extra-pair paternity over the years also lost more within-pair paternity, suggesting a trade-off at the among-individual level. In conclusion, both male and female individual-level phenotypic attributes contribute to male siring success and trade-offs between the different routes may help maintain among-individual phenotypic variance in alternative routes to siring success.

Introduction

Most mating systems offer multiple ways by which males can maximize their siring success (Gross 1996). Naive evolutionary models predict that directional selection will deplete variation in any such siring route (Lynch & Walsh 1998), resulting in the evolution of a 'Darwinian demon' (Law 1979). Natural populations, in contrast, harbor substantial among-individual variation in traits related to siring success (Mousseau & Roff 1987), and phenotypic stasis is commonly observed (Merilä et al. 2001). In socially monogamous animals, extra-pair and within-pair reproduction represent key alternative routes to male siring success (Webster et al. 1995; Griffith et al. 2002). Interestingly, the mechanisms constraining the adaptive evolution of extra-pair and within-pair siring routes have not received much attention. Two key aspects can constrain siring routes in socially monogamous systems; female effects on siring success (Kvarnemo & Simmons 2013; Reid et al. 2014a), and trade-offs between conflicting siring routes (Westneat et al. 1990; Kokko 2005). Quantifying such constraints is challenging in wild populations because trade-offs are often hidden at specific hierarchical levels of phenotypic organization (e.g. within-individuals as opposed to among-individuals: Stearns 1989), while male siring success is simultaneously affected by the phenotype of both members of the social pair. In this paper, we address this issue by using a (co)variance partitioning approach to study alternative routes to siring success in great tits (*Parus major*). Specifically, we estimated the relative contribution of male and female phenotype on male siring success and how these phenotypic effects result in trade-offs between alternative siring routes.

Socially monogamous males can increase their siring success via three routes: mating with highly fecund females, avoiding within-pair paternity loss and seeking extra-pair fertilizations. Total male siring success will be determined by variation in and covariation among these routes (Webster et al. 1995; Lebigre et al. 2013). This (co) variation is in turn determined by the joint effects of male's and his social female's phenotypic characteristics as well as environmental variation (Petrie & Kempenaers 1998; Westneat et al. 2011). For example, if higher resource holding potential increases a male's ability to avoid within-pair paternity loss while simultaneously increasing his chances to gain extra-pair paternity, differences between males in siring success will exist due to among-male variation in resource holding potential. Directional selection is expected to deplete among-male variation in resource holding potential and both routes to siring success, unless trade-offs or other

processes constrain their adaptive evolution. For example, investment in avoidance of within-pair paternity loss might trade-off with investment in obtaining extra-pair copulations in instances where males face limitations in the time or energy available for these activities (Westneat & Stewart 2003; Kokko 2005). These trade-offs would result in a small or zero net evolutionary response to selection of both siring routes, maintaining variation in the siring routes and associated traits. Therefore, the sources of variation in and covariation among siring routes will determine the evolutionary trajectories of male reproductive strategies.

Several studies quantifying the covariance between the different routes to siring success find that males that are successful in avoiding within-pair paternity loss are also more —instead of less—likely to gain extra-pair paternity (Webster et al. 1995; Kempenaers et al. 1997; Schlicht & Kempenaers 2013; Reid et al. 2014b). Such patterns are, notably, not in disagreement with the notion of trade-offs between the siring routes (Noordwijk & Jong 1986). This is because covariances between life-history traits arise due to processes that act at different hierarchical levels. In territorial species, for example, spatial variation in the availability of resources would enable males with 'high-quality' territories to invest in multiple costly activities whereas males with 'low-quality' territories are unable to do so. This would cause a positive covariance at the among-male level (e.g. people with big houses also have big cars; Reznick et al. 2000). Simultaneously, resources invested in one activity cannot be invested in another activity, hence, trade-offs might be revealed only when considering the within-male among-year level (e.g. in the year where people buy a new house they cannot buy a new car). Analogously, the documented positive covariance between siring routes has been interpreted as arising from variation in 'male quality' (Jennions & Petrie 2000), which represents an among-individual process. Higher quality males would be better suited to gain extra-pair paternity, be more successful at avoiding within-pair paternity loss, and be better able to acquire high quality mates (e.g. mates that lay relatively large clutches). At the same time, investment in avoiding within-pair paternity loss might come with less time and/or resources available for investment in extra-pair paternity within a specific breeding attempt. This trade-off would be captured statistically by the sign and magnitude of the within-male among-year covariance in repeated measures datasets. The existence of multiple processes contributing to patterns of covariance thus warrants study designs enabling the partitioning of relationships between siring routes across hierarchical levels.

Quantifying trade-offs among siring routes in socially monogamous animals is more

challenging because each of the routes is difficult to define as the property of a single individual. For example, within-pair paternity loss is often treated as a male trait, although it depends both on the male's ability to secure within-pair fertilizations and the promiscuity level of his social mate (Petrie & Kempenaers 1998; Westneat & Stewart 2003; Reid et al. 2014a). Moreover, female fecundity can affect male within-pair paternity loss because highly fecund females produce more eggs to be fertilized, resulting in higher statistical chances that their male partner loses within-pair paternity. It follows, that the evolutionary dynamics of male siring routes will not only depend on the sources of (co) variation within-sexes but also the sources of covariation across-sexes (Reid et al. 2014a). Female phenotypic characteristics causing variation in and covariation among male siring routes can be viewed as environmental effects on male siring routes, with the particularity that these environmental components have genes and can thus evolve. In the quantitative genetic literature, traits whose expression is affected by the phenotype of other individuals are sometimes called 'interactive phenotypes' (Moore et al. 1997). This type of interaction could result in a source of evolutionary constraint (Wolf et al. 1998; Wolf 2003), because of conflicts of interest between the members of the social pair (Brommer & Rattiste 2008). Variance partitioning approaches (detailed below) have proven very insightful in the study of the sources of variation and covariation of this type of phenotype (McGlothlin & Brodie 2009; Dingemanse & Araya-Ajoy 2014).

We used a variance partitioning approach to quantify the sources and levels of covariation between male siring routes in wild great tits, to understand possible mechanisms constraining their adaptive evolution. Great tits are socially monogamous birds that commonly engage in extra pair reproduction (Brommer et al. 2007; van Oers et al. 2008; Patrick et al. 2012). During four consecutive years (2010-2013), we monitored the breeding ecology of a population of great tits breeding in 12 nest box plots and measured annual male siring success (defined as the total number of eggs sired by a male in each year). We had two main objectives: i) estimate the extent to which male siring success is determined by male and female characteristics and ii) determine whether relationships between different siring routes could result in trade-offs constraining male siring success. Regarding the first objective, we first decomposed male annual siring success into its underlying components: clutch size (the number of eggs produced by his social mate), within-pair paternity loss (number of eggs laid by his social mate that were sired by another male) and extra-pair paternity gain (the number

of eggs that a focal male sired with females other than its social mate) (Fig. 1). We then used a variance partitioning approach to quantify the relative contribution of male and female "identity" effects to variation in male's annual siring success and its three routes. We refer to "identity" effects as the phenotypic characteristics that vary among-individuals (due to genes and/or permanent environmental effects) and cause variation in any of the siring routes. The variance partitioning approach does not provide information about the specific individual-level phenotypes of males or females affecting male's annual siring success, but can be used to quantify overall phenotypic female and/or male effects on siring success that are not caused by phenotypic variation due to plastic responses to short term environmental effects (Griffing 1967).

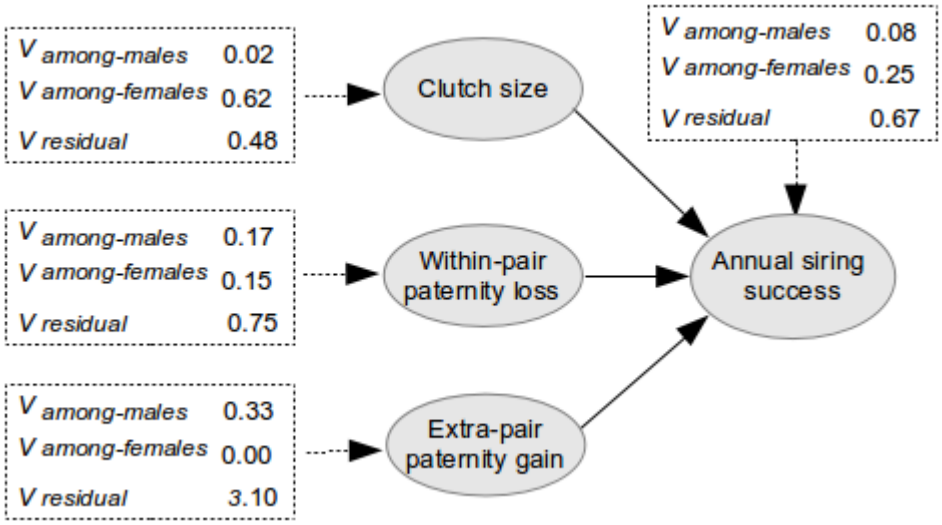

Figure 1. *Path diagram depicting the magnitude of male and female identity effects on annual male siring success for each of three fertilization routes. Boxes with dotted lines show female, male and residual variances (V) estimated from the mixed-effect models shown in table 1.*

To achieve our second objective, we determined how variation in and covariation among siring routes affected male annual siring success. We tested a set of predictions (P) regarding the relations among the different siring routes and male annual siring success (Fig. 2). We logically expected that annual siring success would be affected positively by clutch size (P1 in Fig. 2a), negatively by within-pair paternity loss (P2 in Fig. 2a), and positively by extra-pair paternity gain (P3 in Fig. 2a). We also hypothesized that a bigger clutch should increase a male's probability of losing within-pair paternity. This will result in a trade-off between mating with a highly fecund female and losing within-pair paternity (P4 in Fig. 2a).

We had no reason to expect a direct effect of clutch size on extra-pair paternity gain; therefore we had no prediction about the magnitude or sign of that relationship. For the relationship between extra-pair paternity gain and within-pair paternity loss, we had specific predictions for the within- versus among-male levels. We predicted a negative among-male covariance between extra-pair paternity gain and within-pair paternity loss (P5 in Fig. 2b), assuming that 'high-quality' males (or males possessing high-quality territories) gain more extra-pair paternity while simultaneously being able to avoid within-pair paternity loss (Kempenaers et al. 1997; Jennions & Petrie 2000). In contrast, at the within-male among-year level we expected a trade-off between within- and extra-pair siring routes, statistically expressed as a positive covariance between within-pair paternity loss and extra-pair paternity gain (P5 in Fig. 2c).

As a final step, we investigated the role of a male's behavior in mediating trade-offs between siring routes. Males of the studied population show long-term repeatable differences in how aggressively they respond to simulated territorial intrusions of male conspecifics during their mate's fertile phase (Araya-Ajoy & Dingemanse 2014). Though aggressiveness is repeatable, males are simultaneously plastic, varying their level of aggressiveness within and across years. Since males are only able to lose paternity prior to clutch completion, aggressiveness might thus either directly or indirectly (through correlations with mate guarding intensity) enable males to avoid within-pair paternity loss. We therefore hypothesized that increased aggressiveness towards male intruders during their mate's fertile phase would reduce within-pair paternity loss (P6 in Fig. 2). If investment in securing within-pair fertilizations would indeed trade-off with the expression of behaviors that enables males to gain extra-pair copulations, we also expected a negative effect of aggressiveness on extra-pair paternity gain (P7 in Fig. 2). We further partitioned the effects of aggressiveness on both siring routes into among- versus within-individual levels. We hypothesized that the same mechanism was underlying this relationship at both levels, that is, we expected similar magnitudes and signs of correlations at both levels. We had no a priori reason to predict a relationship between aggressiveness and clutch size, therefore this relationship was not tested.

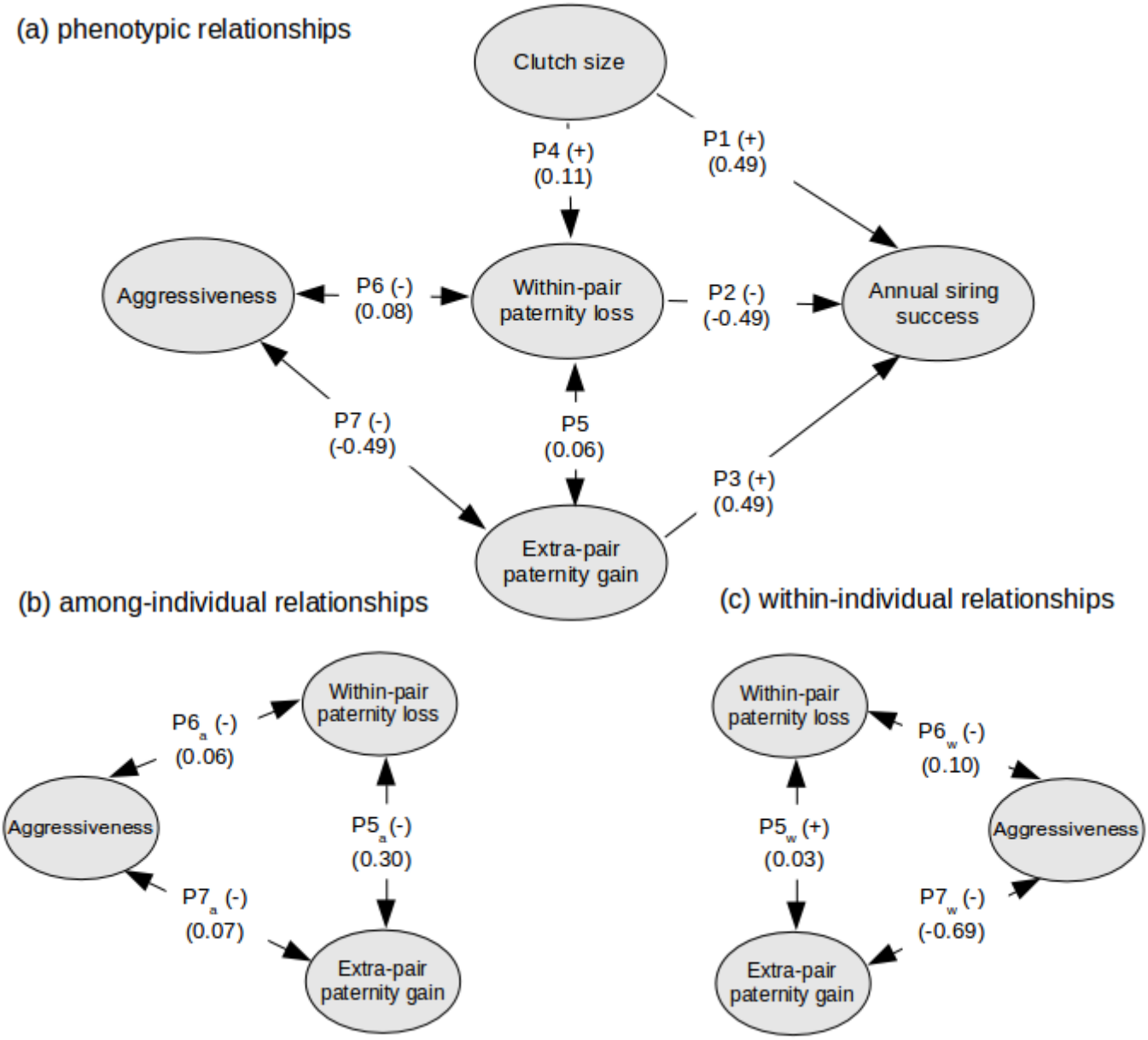

Figure 2. *Path diagram presenting predictions and point estimates derived from our data analyses. Single-headed arrows represent presumed causal relationships (β); double-headed arrows represent correlations between traits (r). Predictions (1-8) are symbolized by the letter P with their direction (- versus +) indicated within parentheses as explained in the main text. When predictions (P) were different across levels we printed the subscript "a" for among-individual relations and the subscript "w" for within-individual relations. In parentheses we also print the mode of the posterior distribution of parameter estimates for the hypothesized effect based on our analyses. The upper panel (a) depicts (unpartitioned) phenotypic relationships between the fertilization routes and male siring success. The lower panels depict relationships between within-pair paternity loss, extra-pair paternity gain and aggressiveness at the among-individual level (b) and at the within-individual among-year level(c).*

Methods

Study site

We studied a population of great tits breeding in 12 nest box plots, established in 2009 in Southern Germany in an area of approximately 120 ha (Bavarian Landkreis Starnberg; 47º 58' N, 11º 14' E). Each plot consists of a regular grid of 50 boxes, with 50 meters between adjacent boxes. From April onwards, boxes were checked twice a week to determine lay date (back-calculated assuming that one egg was laid per day), onset of incubation, and clutch size. When the nestlings were 6 days old, they were blood sampled and marked with an aluminum ring; any unhatched eggs or deceased nestlings were collected. Parents were caught with a spring trap in the nest box the next day, measured, bled, and marked with a unique combination of rings if caught for the first time.

Male variation in siring success

We recorded a total of 6722 eggs in our population distributed in 836 first clutches (nests starting within 30 days after the first egg of the focal year in each plot was found). Because we were interested in male siring success and did not want to bias our measure by variation in hatching success or early survival of within- or extra-pair offspring (García-González 2008), we tried to genotype all successfully fertilized eggs (i.e. hatched nestlings, unhatched eggs and nestlings deceased prior to blood sampling). We were able to genotype 5347 (79%) of the 6752 recorded eggs. We proceeded to perform genetic parentage assignments for these 5347 fertilized eggs using genetic and spatial information incorporated in Bayesian full probability models (R package MasterBayes; Hadfield et al. 2006). We excluded all breeding attempts where maternity was uncertain (genetic mother not sampled, N = 49 broods) and used a 90% confidence cut-off to take a paternity assignment to further analyses. This resulted in 4018 offspring (75 % of the 5,347 genotyped offspring) with assigned paternity from 454 males and 467 female parents in 668 breeding attempts (for further details see Supporting Material Appendix S1; and Table S1 for a description of the markers used). Given current debates on the pros and cons of alternative paternity assignment methods (Walling et al. 2010), we also performed the paternity assignment in another commonly used package (Cervus 3.0.6); this produced similar findings (results not shown). We estimated male within-pair paternity loss as the number of eggs produced by the social female of a nest that were not sired by the social male, and male extra-pair paternity gain as the number of eggs

that a focal male sired with females other than its social mate. Male annual siring success was approximated as the sum of clutch size and extra-pair paternity gain minus within-pair paternity loss.

Male aggressiveness assay

We measured male aggressive responses to standardized simulated territorial intrusions for each first brood. Each male was subjected to two aggression tests during the fertile period of its social mate (one and three days after the first egg was found). The behavioral test started when a taxidermic mount of a male great tit with a playback song was presented one meter away from the subject's nest box on a 1.2 meter wooden pole. We subsequently recorded the behavior of the focal male for a period of three minutes after it had entered a 15-meter radius around the nest box. Details of the experimental setup, and assayed behaviors, are given in Araya-Ajoy & Dingemanse (2014). As a measure of the intensity of the aggressive response, we used the subject's minimum approach distance during the assay. We obtained 784 measures of aggressiveness distributed in 516 (78% of the 668) breeding attempts of 386 (85% of the 454) males.

Statistical analyses

Variance partitioning of routes to male siring success

We first quantified the sources of variation in male annual siring success (Model 1) and its three underlying components ('routes'): clutch size (Model 2), within-pair paternity loss (Model 3) and extra-pair paternity gain (Model 4). We used mixed-effect models to determine variance attributable to male identity (n=454 individual males), female identity (n=467 individual females) versus unidentified exogenous variables (i.e. residual variance; 668 observations). We were able to disentangle male versus female identity effects because we had repeated measure across years for 158 out of 454 (35%) males (no. of individuals (no. of years): 294 (1) 114 (2), 38 (3), 8 (4)), 149 out of 467 (31%) females (321 (1), 101 (2), 35 (3), 10 (4)), and because 160 out of 454 (35%) males bred with different females across years (294 (1), 114 (2), 38 (3), 8 (4)). To achieve the variance partitioning, we used mixed-effect models with random intercepts for male and female identity. For these and all subsequent models, annual siring success and clutch size were standardized to a variance of one and were modeled with Gaussian errors. Within-pair paternity loss and extra-pair paternity gain were not transformed and modeled assuming a Poisson over-dispersed distribution (for

biological reasons, see (Brommer et al. 2007)). Over-dispersion was modeled by including an observation level random effect which may alleviate zero-inflation problems (Hinde & Demétrio 1998). Extra-pair paternity gain and within-pair paternity loss are often modeled as binomial responses, therefore we provided comparable results in Table S2 (Appendix S2).

Relationships between routes to male siring success

We estimated predicted relationships between fertilization routes with a series of mixed-effect models. This allowed us to quantify presumed causal effects as fixed-effect estimates (e.g. the effect of clutch size on within-pair paternity loss) as well as (co)variances associated with random effects at the among- and within-individual male level (e.g. the covariance between paternity loss and gain). We first used a mixed-effect model to determine the contribution of extra-pair paternity gain, within-pair paternity loss and clutch size to a male's annual siring success (Fig. 2, relationships P1-P3). We then used a mixed-effect model to determine the relation between clutch size and within-pair paternity loss (Fig. 2, P4). This model had random intercepts for male identity and assumed an over dispersed-Poisson error distribution. Afterwards, we performed a bivariate mixed-effect model to study the relation between extra-pair paternity gain and within-pair paternity loss at the different levels (Table 2), which allowed us to quantify prediction $P5_a$ in Figure 2b and $P5_w$ in Figure 2c. The model consisted of two response variables: extra-pair paternity gain and within-pair paternity loss. We included random intercepts for male identity and calculated the among-male and residual (within-individual) covariances between these two siring routes as well as the overall (un-partitioned) phenotypic correlation (Wilson et al. 2010; Dingemanse & Dochtermann 2013). We did not include random intercepts for female identity to avoid over-parameterization, therefore variance and covariances in the response variables associated with this effect will be captured in the within-male residual covariance (Wilson et al. 2010).

We then performed two bivariate models to determine the covariance between male aggressiveness and within-pair paternity loss and extra-pair paternity gain, all fitted as response variables. We fitted random intercepts for male and breeding attempt identity. This enabled us to quantify the covariance between the response variables within-individuals among-years (i.e. environmental covariance) versus among-individuals (i.e. long-term (cross-year) individual differences) while fully acknowledging the replicated structure of our sampling design. Because aggressiveness was assayed multiple times per year, variation in

aggressiveness was partitioned into variance attributable to male identity, breeding attempt identity, and residual (within-breeding attempt) variance. In contrast, because each individual had a single value per year for within-pair paternity loss and extra-pair paternity gain, variation in these response variables was partitioned into male and breeding attempt identity effects while constraining the residual variance (and covariance) to zero. We assumed over-dispersed Poisson error distributions for both variables. Given that lower values of minimum approach distance reflect higher aggressiveness, for the results section we multiplied the covariance between the male minimum approach distance and the routes to siring success by -1 for ease of interpretation.

General modeling procedures

Data manipulation and statistical analyses were conducted in R v 3.1. Mixed-effect models were fitted using restricted maximum likelihood algorithms in the lme4 package (Bates *et al.* 2014). Posterior distributions were simulated using the sim function of the arm package (Gelman & Hill 2007) to determine the uncertainty around model estimates. We present mode (as a measure of central tendency) and the 95% credible intervals (as a measure of uncertainty of the posterior distribution) of parameter estimates. A full description of how bivariate models were fitted in the lme4 package and a simulation study to assess the robustness of this approach is given in Appendix S3. To corroborate the results of the bivariate mixed-effect models we ran the analyses in a Bayesian framework using JAGS (Plummer 2003) within R, which produced qualitatively similar results (results not shown).

Results

Male paternity: descriptive statistics

Clutch size ranged from 3-13 eggs with a mean of 8.14. Mean within-pair paternity loss was 0.62 and ranged from 0-6 offspring, with 37% of males losing at least one offspring. Mean extra-pair paternity gain was 0.43 and ranged from 0-8 offspring, with 23% of males siring at least one extra-pair offspring. Mean siring success was 7.86, and ranged from 2-16 offspring. In a closed population the amount of gain and loss should be the same, because our population was not closed and we could not assigned the extra-pair fathers to all of the chicks (some extra-pair sires were breeding outside of the population or were floaters) the population-average estimates of extra-pair gain and within-pair loss are somewhat different.

Variance components of annual siring success and its routes

Analysis of the sources of variation in male siring success demonstrated that (unidentified) individual-specific traits of females contributed more strongly to male siring success than individual-specific traits of males (variance attributable to female identity: 25%, to male identity: 8%), while most of the variation remained unexplained (67%, Table 1). The relative contribution of male versus female identity effects greatly differed between siring routes (Table 1; Fig. 1): individual-specific traits of males explained significantly more variation in extra-pair paternity gain than female individual-specific traits, whereas variance in clutch size was largely attributable to female identity effects (62%) (Table 1). Finally, both male (17%) and female (15%) identity effects explained similar variance in within-pair paternity loss (Table 1).

Table 1. Variance components (*V*) for annual male siring success and its different components. Estimates presented are the mode of the posterior distribution and in parentheses the lower and upper credible interval limits (95% CI).

Response variable	Male annual siring success	Clutch size	Within-pair * paternity loss	Extra-pair * paternity gain
Fixed effects	β (95% CI)	β (95% CI)	β (95% CI)	β (95% CI)
Intercept	-0.01 (-0.09, 0.08)	0.00 (-0.09, 0.08)	-0.96 (-1.11, -0.83)	-2.14 (-2.41, -1.93)
Random effects	σ^2 (95% CI)	σ^2 (95% CI)	σ^2 (95% CI)	σ^2 (95% CI)
$V_{among\text{-}males}$ n= 454 males	0.08 (0.07, 0.10)	0.02 (0.01, 0.02)	0.17 (0.16, 0.20)	0.33 (0.29, 0.38)
$V_{among\text{-}females}$ n=467 females	0.25 (0.20, 0.27)	0.62 (0.57, 0.74)	0.15 (0.12, 0.16)	0.00 -
$V_{among\text{-}observations}$ n=668 breeding attempts	0.67 (0.62, 0.77)	0.48 (0.43, 0.53)	0.75 (0.66, 0.82)	3.10 (2.71, 3.35)

**Untransformed estimates from a generalized mixed-effect models with Poisson error distribution.*

Relationships between siring routes

In our path model, variation in annual siring success was by definition positively affected by extra-pair paternity gain and clutch size, and negatively affected by within-pair paternity loss (Fig 2a). These three routes, by definition, accounted for all the variation in male annual siring success (Appendix S2, Table S3).

In line with our predictions (P4, Fig. 2), we found that males with large clutches also

lost more within-pair paternity (mode=0.11; 95% CI=0.03 – 0.20; Appendix S2, Table S4). The relationship between extra-pair paternity gain and within-pair paternity loss was close to zero at the within-male (among-year) level (Table 2; Fig. 2c, P5). In contrast, the correlation was significantly positive at the among-male level (Table 2; Fig 2b, P5). In other words, the data were consistent with the presence of a trade-off between within-pair paternity loss and extra-pair paternity gain, but not at the within-individual level where it was expected.

Table 2. Results of a bivariate mixed-effect model used for estimating the relationship between routes to male fertilization success. Random effects estimates are variances (*V*), covariances (*COV*) and correlations (*r*). Estimates presented are the mode of the posterior distribution and in parentheses the lower and higher credible interval (95% CI).

	Within-pair* paternity loss	Extra-pair* paternity gain
Fixed effects	β (95% CI)	β (95% CI)
Intercept	-0.95 (-1.08, -0.81)	-2.12 (-2.36, -1.88)
Random effects	σ^2 (95% CI)	σ^2 (95% CI)
Among-males (n=454)		
$V_{male\ ID}$	0.17 (0.15, 0.19)	0.33 (0.29, 0.38)
$Cov_{gain\text{-}loss}$	0.07 (0.05, 0.09)	
$r_{gain\text{-}loss}$	0.30 (0.22, 0.38)	
Among-broods (n=668)		
V_{broods}	0.85 (0.76, 0.94)	2.77 (2.51, 3.07)
$Cov_{gain\text{-}loss}$	0.04 (-0.07, 0.15)	
$r_{gain\text{-}loss}$	0.03 (-0.05, 0.01)	
Phenotypic correlation		
$r_{gain\text{-}loss}$	0.06 (0.00, 0.12)	

*Untransformed estimates from a generalized mixed-effect models with over-dispersed Poisson error distribution.

Aggressiveness mediating paternity trade-offs?

We tested whether aggressiveness represented the behavioral mediator of the trade-off

between within-pair paternity loss and extra-pair paternity gain, with aggressive animals enjoying decreased within-pair paternity loss at the expense of decreased ability to invest in extra-pair paternity gain. Contrary to our expectations, a male's aggressiveness during its mate's laying phase did not appear to reduce his chances of losing within-pair paternity at any hierarchical level (Table 3a). Instead, in years in which a male showed higher aggressiveness, he also lost more, instead of less, within-pair paternity. On the other hand, in agreement with expectations, in years in which a male was more aggressive he also gained less extra-pair paternity (Table 3b). At the among-male level, aggressiveness did not correlate with extra-pair paternity gain or loss (Table 3b).

Table 3. Results of bivariate mixed-effect models used for estimating the relationship between aggressiveness and (a) within-pair paternity loss or (b) extra-pair paternity gain. Random effects estimates are variances (*V*), covariances (*COV*) and correlations (*r*). We present the mode of the posterior distribution and in parentheses credible intervals(95% CI).

	(a) Aggressiveness*	Within-pair* paternity loss	(b) Aggressiveness*	Extra-pair* paternity gain
Fixed effects	β (95% CI)	β (95% CI)	β (95% CI)	β (95% CI)
Intercept	1.69 (1.60, 1.77)	-0.87 (-1.02, -0.73)	1.68 (1.59, 1.78)	-2.04 (-2.28, -1.81)
Random effects	σ^2 (95% CI)	σ^2 (95% CI)	σ^2 (95% CI)	σ^2 (95% CI)
Among-males *(n=386)*				
$V_{male\ ID}$	0.26 (0.23, 0.30)	0.20 (0.17, 0.23)	0.31 (0.26, 0.34)	0.18 (0.16, 0.22)
$Cov_{gain\text{-}loss}$	0.00 (-0.02, 0.03)		0.02 (-0.02, 0.03)	
$r_{gain\text{-}loss}$	0.06 (-0.03, 0.14)		0.07 (-0.06, 0.13)	
Among-broods *(n=516)*				
V_{broods}	0.24 (0.28, 0.35)	0.60 (0.59, 0.74)	0.40 (0.36, 0.45)	2.26 (2.05, 2.55)
$Cov_{gain\text{-}loss}$	0.04 (0.01, 0.08)		-0.64 (-0.75, -0.57)	
$r_{gain\text{-}loss}$	0.10 (0.04, 0.14)		-0.69 (-0.72, -0.64)	
Phenotypic correlation				
$r_{gain\text{-}loss}$	0.08 (0.02, 0.14)		-0.49 (-0.53, -0.44)	
Residual				
$V_{residual}$	0.59 (0.54, 0.65)		0.57 (0.52, 0.63)	

Untransformed estimates from a generalized mixed-effect model with over-dispersed Poisson error distribution.

Discussion

We quantified the sources of variation in male siring success in wild great tits in four consecutive years to determine whether males alternative siring routes were constrained by i) the phenotype of his social mate and ii) trade-offs between siring routes. Females substantially influenced male siring success, particularly via clutch size, but also via within-pair paternity loss (Fig 1, Table 1). Those effects were mediated by repeatable variation in female fecundity and female promiscuity (Table 1). Moreover, females producing large clutches had more eggs sired by an extra pair male (Fig 2a). This pattern is consistent with a trade-off between mating with a highly fecund female and avoiding within-pair paternity loss. Males that consistently gained more extra-pair paternity compared to others also lost more within-pair paternity, a finding that is consistent with an among-male trade-off between these siring routes (Fig 2b). In conclusion, the patterns of (co)variation among siring routes suggest that female phenotypic effects and both inter- and intra-sex trade-offs constrain the adaptive evolution of male alternative siring routes. When studying the role of aggressiveness in mediating the investment in the different siring routes, we found that aggressiveness did correlate with both extra-pair paternity gain and within paternity loss (Figure 3, Table 3), but not in a way that suggesting that aggressiveness is the behavioral mediator of the expected within-male (or the detected among-male) paternity trade-off.

Female phenotype affects male siring success

Female identity explained two times more variance in male siring success compared to male identity (Table 1). If female and male identity effects represent genetic effects, natural selection will mainly act on male alternative routes to siring success via heritable variation in females (an indirect genetic effec; Wolf et al. 1998). Our analyses suggest that female effects acted particularly through clutch size, i.e. due to among-female variation in fecundity. In contrast, male identity explained a minor portion (2%) of the variation in clutch size. This result suggests that differential female investment in response to repeatable among-male variation in phenotypic attributes (i.e. male quality) plays a minor role, which is consistent with other studies in Great tits (Browne et al. 2007). Female identity also explained (some) variation (~14%) in within-pair paternity loss of her social mate. Repeatable among-female variation in promiscuity will inherently lead to such female identity effects (see Reid et al. 2014a). Male within-pair paternity loss is the same trait as his social mate's extra-pair

paternity reproduction. For this reason, extra-pair paternity has been referred to as a "meta-trait" (Westneat & Stewart 2003), given that it is determined by at least three players: the cuckolded male, the extra-pair father and the female (see also Box 1 in Petrie & Kempenaers 1998). Female effects on male siring routes can thus be viewed as a social environment effect on male siring success. Indirect genetic effects theory developed in quantitative genetics implies that such social environment effects may impose major evolutionary constraints (Wolf et al. 1998; McGlothlin et al. 2010; see Brommer & Rattiste (2008) for an empirical example). The reported female effects may thereby help explain why phenotypic stasis, as well as among-male variation in the different siring routes might persist despite the expected directional selection.

From the female's perspective, benefits of extra-pair reproduction remain generally not clear (Forstmeier et al. 2014). If females engage in extra-pair reproduction to reap genetic benefits (Jennions & Petrie 2000), this would cause an inter-sexual conflict between the male's efforts to reduce within-pair paternity loss and the female's benefits arising from extra-pair reproduction. Although researchers have been aware that the phenotype of the female mate affects male siring success, this is one of the few studies that empirically quantified the different ways by which such female identity effects can affect the evolution of male siring routes.

Patterns of (co)variation between paternity gain and loss

Males showed substantial among-year repeatability in extra-pair paternity gain and within-pair paternity loss (Table 1, Fig 1), implying that males are predictable in their realized avoidance of paternity loss and extra-pair paternity gain. The long-term (repeatable) parts of these two siring routes were, furthermore, positively correlated (owing to significant among-male covariance). Specifically, males that consistently gained more extra-pair paternity also consistently lost more within-pair paternity. Based on the notion that variation in male resources (Noordwijk & Jong 1986) typically leads to "winners" and "losers", we had expected that males that lose less within-pair paternity would also gain more extra-pair paternity. Our finding is instead consistent with a trade-off between these two siring routes at the among-male level. This positive among-male correlation might be proximately underpinned by a genetic covariance between the two routes (a genetic trade-off) due to opposing pleiotropic effects of genes. An alternative, or complementary, explanation is that individuals specialize in

a specific social niche (Dall et al. 2012; Montiglio et al. 2013): some males might specialize in maximizing within-pair paternity ('defender strategy'), while others might instead specialize in maximizing extra-pair paternity ('intruder strategy'). Importantly, the among-male trade-off that we estimated would not have been detected if we had not used the appropriate variance partitioning approaches, because its effects were hidden due to the modest repeatabilities of paternity loss and gain (Dingemanse et al. 2012; Brommer 2013).

We expected that the patterns of covariance between within-pair paternity loss and extra-pair gain would be driven by different processes at each level. The correlation between within-pair paternity loss and extra-pair paternity gain differed, as expected, significantly between levels (their credible intervals did not overlap; Table 2), although not in the way we had predicted based on life history theory (Noordwijk & Jong 1986). We predicted that the covariance at the among-male level was primarily caused by variation in resources (territory or male "quality"), whereas, the within-individual covariance was caused by a resource allocation trade-off (Noordwijk & Jong 1986). We instead detected a positive covariance at the among-male level suggesting the existence of an among-male trade-off, and no evidence of the predicted within-individual trade-off. Additional mechanisms must be invoked to explain why average level of extra-pair paternity gain and within-pair paternity loss were correlated across years, but within-individual among-year changes were not. One explanation is that within-individual resource allocation trade-offs simply do not exist. This would imply that if an individual during one year decides to invest more in one route, this will not constrain its investment in the other route in the same year. Another possibility is that endogenous or exogenous variables that change within-individuals from one year to another affect these routes differently. For example it has been shown that a male's ability to gain extra-pair paternity increases with age but not its ability to avoid within-pair paternity loss (Cleasby & Nakagawa 2012). Indeed, we did detect this particular pattern also in our study population (Appendix S2, Table S5). Differential effects of age on extra-pair paternity gain versus within-pair paternity loss, might consequently result in a lack of correlation between these routes at the within-individual level, even if a within-individual trade-off exists for other reasons. Another possible explanation of the lack of within-individual correlation is that year to year variation in the environment, like in local density, affects a male's ability to defend its mate but not its ability to gain extra-pair paternity, also obscuring potential within-individual trade-offs.

Aggressiveness and its role in male siring success

We hypothesized that the trade-off between gaining extra-pair paternity and avoiding within-pair paternity loss might be mediated by male aggressive (mate or territory defense) behavior. We expected that investment in aggressive behavior would reduce within-pair paternity loss but will also constrain a male's ability to gain extra-pair paternity due to time or energy allocation trade-offs. We found no evidence for such dual effects. Instead, in years where individuals were more aggressive they also gained less extra-pair paternity, suggesting that investment in aggressive behaviors may indeed constrain their ability to gain paternity. However, investment in aggressive behaviors was also associated with more, not less, within-pair paternity loss (Table 3). This does not necessarily imply that aggressiveness does not prevent within-pair paternity loss, but the observed correlation could instead be the result of the "best-of-a-bad-job" (Kempenaers et al. 1995): paternity loss could have been even higher if those males would not have been aggressive. For example, variation in population density could make individuals behave more aggressive because of an increase threat but they might then also lose more paternity because their increased aggressiveness did not fully secure against paternity loss in such social environments. Modeling the relation between aggressiveness and within-pair paternity loss as a covariance allowed us to view aggressiveness either as a cause or a consequence of patterns in siring. Our results suggest that within-male among-year variation in aggressiveness could be a consequence of within-individual among-year variation in the risk of losing within-pair paternity. Moreover, the overall correlation patterns of aggressiveness with extra-pair paternity gain and within-pair paternity loss, are consistent with the traditional idea that males assess their relative "quality" compared to their social environment (other males) and are aggressive when they have a higher risk of losing paternity and little chance to gain paternity (Kempenaers et al. 1995).

Multi-level covariation and evolutionary responses

Partitioning the levels of covariation between behavior and traits closely related to fitness is key to understanding responses to selection (Roff 1992). Covariation between behaviors and traits closely linked to fitness (such as within pair paternity loss and extra-pair paternity gain) could be due to environmental pleiotropy (i.e. a correlated plastic response of both traits to the same environmental gradient). Importantly, this type of covariation would not result in phenotypic evolution due to natural selection (Sheldon et al. 2003). Responses to

selection are only expected if the covariance between behaviors and fitness are at the among-individual level and underpinned by an additive genetic covariance (Lynch & Walsh 1998). Given that behaviors have an average repeatability of ~ 0.3 (Bell et al. 2009) and on average only 50% of this variation is due to additive genetic effects (Dochtermann et al. 2015), most of the phenotypic correlations between behaviors are not expected to result in an adaptive evolutionary response to selection. Moreover, given the modest repeatabilities of within-pair paternity loss and extra-pair paternity gain, the reported correlations in the literature between these traits and behaviors (Duckworth 2006; Patrick et al. 2012), statistically have largely a within-individual signature (Dingemanse et al. 2012). Indeed, also in our study, there was a significant correlation between aggressiveness and extra-pair paternity gain. This phenotypic correlation was, importantly, mainly due to this within-individual correlation. It is worth noting that if we had not partitioned this correlation into it's among- versus within-individual effects, we could have erroneously concluded that aggressive "personalities" were selected against in our population.

Concluding remarks

This study suggests that both inter- and intra-sex trade-offs exist between routes to male siring success in a species with a socially monogamous mating system. Our study highlights that different biological mechanisms might act across hierarchical levels and that the social environment is an important source of variation in male siring success. In conclusion, acknowledging the notion that siring routes can vary and covary at multiple levels and the particularities of the social environment as source of phenotypic variation is necessary to further our understanding of the evolution of the reproductive strategies of genetically promiscuous species.

Acknowledgments

We thank Anne-Lise Olsen, Erica Stuber, Maria Moiron, Robin Abbey-Lee and all field assistants and students for help in data collection and insightful discussion. Members of the Department of Behavioral Ecology and Evolutionary Genetics, and the Evolutionary Ecology

of Variation Group are thanked for stimulating feedback. Lars Penke, Dave Westneat, Francesca Santostefano and Jon Wright are gratefully acknowledged for feedback, and Renate Heckel-Merz for help with the taxidermic mounts. The study was supported by the German Academic Exchange Service (DAAD), the Max Planck Society, and the International Max Planck Research School for Organismal Biology.

References

Araya-Ajoy, Y. G. & Dingemanse, N. J. 2014. Characterizing behavioural “characters”: an evolutionary framework. *Proceedings. Biological sciences / The Royal Society*, **281**, 20132645.

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

Brommer, J. E. 2013. On between-individual and residual (co)variances in the study of animal personality: are you willing to take the “individual gambit”? *Behavioral Ecology and Sociobiology*, **67**, 1027–1032.

Brommer, J. E. & Rattiste, K. 2008. “Hidden” reproductive conflict between mates in a wild bird population. *Evolution; international journal of organic evolution*, **62**, 2326–2333.

Brommer, J. E., Korsten, P., Bouwman, K. M., Berg, M. L. & Komdeur, J. 2007. Is extrapair mating random? On the probability distribution of extrapair young in avian broods. *Behavioral Ecology*, **18**, 895–904.

Browne, W. J., Mccleery, R. H., Sheldon, B. C. & Pettifor, R. A. 2007. Using cross-clasiffied multivariate mixed response models with application to life history traits in great tits (Parus major). *Statistical Modelling*, **7**, 217–238.

Cleasby, I. & Nakagawa, S. 2012. The influence of male age on within - pair and extra - pair paternity in passerines. *Ibis*, **154**, 318–324.

Dall, S. R. X., Bell, A. M., Bolnick, D. I. & Ratnieks, F. L. W. 2012. An evolutionary ecology of individual differences. *Ecology letters*, **15**, 1189–1198.

Dingemanse, N. J. & Araya-Ajoy, Y. G. 2014. Interacting personalities: behavioural ecology meets quantitative genetics. *Trends in Ecology & Evolution*, **30**, 88–97.

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

Dingemanse, N. J., Dochtermann, N. a. & Nakagawa, S. 2012. Defining behavioural syndromes and the role of “syndrome deviation” in understanding their evolution. *Behavioral*

Ecology and Sociobiology, **66**, 1543–1548.

Dochtermann, N. A, Schwab, T. & Sih, A. 2015. The contribution of additive genetic variation to personality variation: heritability of personality. *Proceedings. Biological sciences / The Royal Society*, **282**, 20142201.

Forstmeier, W., Nakagawa, S., Griffith, S. C. & Kempenaers, B. 2014. Female extra-pair mating: adaptation or genetic constraint? *Trends in Ecology & Evolution*, **29**, 456–464.

García-González, F. 2008. Male genetic quality and the inequality between paternity success and fertilization success: consequences for studies of sperm competition and the evolution of. *Evolution; international journal of organic evolution*, **62**, 1653–1665.

Griffing, B. 1967. Selection in reference to biological groups I. Individual and group selection applied to populations of unordered groups. *Australian Journal of Biological Sciences*, 127–139.

Griffith, S., Owens, I. & Thuman, K. 2002. Extra pair paternity in birds: a review of interspecific variation and adaptive function. *Molecular Ecology*, **11**, 2195–2212.

Gross, M. 1996. Alternative reproductive strategies and tactics: diversity within sexes. *Trends in Ecology & Evolution*, **11**, 92–97.

Hadfield, J. D., Richardson, D. S. & Burke, T. 2006. Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework. *Molecular ecology*, **15**, 3715–3730.

Hinde, J. & Demétrio, C. G. B. 1998. Overdispersion: Models and estimation. *Computational Statistics & Data Analysis*, **27**, 151–170.

Jennions, M. D. & Petrie, M. 2000. Why do females mate multiply? A review of the genetic benefits. *Biological reviews of the Cambridge Philosophical Society*, **75**, 21–64.

Kempenaers, B., Verheyen, G. R. & Dhondt, A. A. 1995. Mate guarding and copulation behaviour in monogamous and polygynous blue tits: do males follow a best-of-a-bad-job strategy? *Behavioral Ecology and Sociobiology*, **36**, 33–42.

Kempenaers, B., Verheyen, G. R. & Dhondt, A. A. 1997. Extrapair paternity in the blue tit (Parus caeruleus): female choice, male charateristics, and offspring quality. *Behavioral Ecology*, **8**, 481–492.

Kokko, H. 2005. Mate guarding, male attractiveness, and paternity under social monogamy. *Behavioral Ecology*, **16**, 724–731.

Kvarnemo, C. & Simmons, L. 2013. Polyandry as a mediator of sexual selection before and after mating. *Philosophical transactions of the Royal Society of London. Series B, Biological*

sciences, **368**, 20120042.

Law, R. 1979. Optimal life histories under age-specific predation. *American Naturalist*, **114**, 399–417.

Lebigre, C., Arcese, P. & Reid, J. M. 2013. Decomposing variation in male reproductive success: age-specific variances and covariances through extra-pair and within-pair reproduction. *The Journal of Animal Ecology*, **82**, 872–883.

Lynch, M. & Walsh, B. 1998. *Genetics and analysis of quantitative traits.* Sunderland, Mass.: Sinauer.

McGlothlin, J. W. & Brodie, E. D. 2009. How to measure indirect genetic effects: the congruence of trait-based and variance-partitioning approaches. *Evolution; international journal of organic evolution*, **63**, 1785–1795.

McGlothlin, J. W., Moore, A. J., Wolf, J. B. & Brodie, E. D. 2010. Interacting phenotypes and the evolutionary process. III. Social evolution. *Evolution; international journal of organic evolution*, **64**, 2558–2574.

Merilä, J., Sheldon, B. & Kruuk, L. 2001. Explaining stasis: microevolutionary studies in natural populations. *Genetica*, **112-113**, 199–222.

Montiglio, P., Ferrari, C. & Réale, D. 2013. Social niche specialization under constraints: personality, social interactions and environmental heterogeneity. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences*, **368**, 20120343.

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

Mousseau, T. & Roff, D. 1987. Natural selection and the heritability of fitness components. *Heredity*, **59**, 181–197.

Noordwijk, A. Van & Jong, G. De. 1986. Acquisition and allocation of resources: their influence on variation in life history tactics. *American Naturalist*, **128**, 137–142.

Patrick, S. C., Chapman, J. R., Dugdale, H. L., Quinn, J. L. & Sheldon, B. C. 2012. Promiscuity, paternity and personality in the great tit. *Proceedings. Biological sciences / The Royal Society*, **279**, 1724–1730.

Petrie, M. & Kempenaers, B. 1998. Extra-pair paternity in birds: explaining variation between species and populations. *Trends in Ecology & Evolution*, **13**, 52–58.

Reid, J. M., Arcese, P., Keller, L. F. & Losdat, S. 2014a. Female and Male Genetic Effects on Offspring Paternity: Additive Genetic (Co)Variances in Female Extra-Pair Reproduction

and Male Paternity Success in Song Sparrows (Melospiza Melodia). *Evolution; international journal of organic evolution*, 1–14.

Reid, J. M., Arcese, P. & Losdat, S. 2014b. Genetic covariance between components of male reproductive success: within-pair vs. extra-pair paternity in song sparrows. *Journal of evolutionary biology*, **27**, 2046–2056.

Reznick, D., Nunney, L. & Tessier, A. 2000. Big houses, big cars, superfleas and the costs of reproduction. *Trends in Ecology & Evolution*, 421–425.

Roff, D. A. 1992. *Evolution Of Life Histories: Theory and Analysis.* Springer.

Schlicht, E. & Kempenaers, B. 2013. Effects of social and extra-pair mating on sexual selection in blue tits (Cyanistes caeruleus). *Evolution*, **67**, 1420–1434.

Sheldon, B. C., Kruuk, L. E. B. & Merila, J. 2003. When environmental variation short-circuits natural selection. **18**, 207–209.

Stearns, S. 1989. Trade-offs in life-history evolution. *Functional ecology*, **3**, 259–268.

Van Oers, K., Drent, P. J., Dingemanse, N. J. & Kempenaers, B. 2008. Personality is associated with extrapair paternity in great tits, Parus major. *Animal Behaviour*, **76**, 555–563.

Walling, C. A, Pemberton, J. M., Hadfield, J. D. & Kruuk, L. E. B. 2010. Comparing parentage inference software: reanalysis of a red deer pedigree. *Molecular ecology*, **19**, 1914–1928.

Webster, M., Pruett-Jones, S., Westneat, D. & Arnold, S. 1995. Measuring the effects of pairing success, extra-pair copulations and mate quality on the opportunity for sexual selection. *Evolution; international journal of organic evolution*, **49**, 1147–1157.

Westneat, D. F. & Stewart, I. R. K. 2003. Extra-pair Paternity in Birds: Causes, Correlates, and Conflict. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 365–396.

Westneat, D., Sherman, P. & Morton, M. 1990. The ecology and evolution of extra-pair copulations in birds. *Current Ornithology*,

Westneat, D. F., Hatch, M. I., Wetzel, D. P. & Ensminger, A. L. 2011. Individual variation in parental care reaction norms: integration of personality and plasticity. *The American naturalist*, **178**, 652–67.

Wilson, A. J., Réale, D., Clements, M. N., Morrissey, M. M., Postma, E., Walling, C. a, Kruuk, L. E. B. & Nussey, D. H. 2010. An ecologist's guide to the animal model. *The Journal of Animal Ecology*, **79**, 13–26.

Wolf, J. B. 2003. Genetic architecture and evolutionary constraint when the environment contains genes. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 4655–4660.

Wolf, J. B., Brodie, E. D., Cheverud, J. M., Moore, A. J. & Wade, M. J. 1998. Evolutionary consequences of indirect genetic effects. *Trends in Ecology & Evolution*, **13**, 64–69.

Appendix S1. Paternity assignment.

DNA sampling

Each study year (2010 through 2013), a small blood sample (~5 µL) was taken from the brachial vein of all ringed chicks and adults. Unhatched eggs and dead chicks were collected during nest visits at chick age day 6 and d14, and we tried to collect any remaining unhatched eggs and dead chicks when the breeding attempt finished (either when the chicks fledged or if the parents abandoned the nest). We obtained blood samples from 984 out of 1099 breeding attempts (90%). Analyses presented in the main text are only focusing on first clutches (668 out of the 984).

DNA extraction and microsatellite genotyping

Genomic DNA was extracted from blood samples stored in Queen's Lysis buffer (0.01 M Tris-HCl, 0.01 M NaCl, 0.01 M Na-EDTA, 1% n-Lauroylsarcosine, pH 8.0; Seutin et al. 1991) using the NucleoSpin Blood Quick Pure Kit (Macherey-Nagel GmbH, Düren, Germany). The DNeasy Blood and Tissue Kit (QIAGEN) was used to isolate DNA samples from dead chicks and unhatched embryos of large or medium size. DNA from germinal disks of unhatched eggs was extracted with a standard Phenol-Chloroform protocol (Sambrock & Russel 2001).

Twenty three variable polymorphic microsatellite markers and one sex chromosome linked marker were used for genotyping (Table S5): Pca7, Pca8, Pca9 (Dawson et al. 2000); POCC6 (Bensch et al. 1997); Mcy□4 (Griffiths & Double 1998); PmaC25, PmaGAn27, PmaTGAn33, PmaTGAn42, PmaTAGAn71, PmaTAGAn86, PmaD105, PmaD130 (Saladin et al. 2003); ClkpolyQ (Johnsen et al 2007); ADCbm, NPAS2 (Steinmeyer et al. 2010); DRD4 UTR Indel, DRD4 ID13606 (unpublished, Jakob Mueller, Christine Baumgartner); Titgata68 (Wang et al. 2005); CcaTgu6, CcaTgu19, CcaTgu27 (Olano-Marin et al. 2010); DkiB102-ZEST (primer development Olano-Marin et al., 2010); original sequence isolation (King et al. 2005) and the sex chromosome linked P2/P8 (Griffiths & Double 1998).

Microsatellite amplifications were performed in multiplexed PCRs using the Qiagen Type-it Microsatellite PCR Kit (Qiagen, Hilden, Germany) and primer mixes containing six or seven primer pairs (mix 1A – 3A in 2010 to 2012 and mix 1B - 3B in 2013), Table 1). The forward primer of each pair was fluorescently labeled with 6-FAM, VIC, PET or NED (Dye Set G5; Applied Biosystems, Darmstadt, Germany). Differences in amplification efficiency and dye strength of the primers were accommodated by adapting the primer concentrations in these mixes (details given in Table 1). Each 10 µl multiplex PCR contained 20 – 80 ng DNA, 5 µl of the 2x Type-it Microsatellite PCR Master Mix and 1 µl of one of a primer mix. Cycling conditions were: 15 min initial denaturation at 95 ˚C, 28 cycles of 30 s denaturation at 94 ˚C, 90 s annealing at the temperature given in Table1, and 1 min extension at 72 ˚C, followed by 30 min completing final extension at 60 ˚C. After amplification, 1.5 µl of the PCR products were added to 13ul formamide containing the GeneScan 500 LIZ Size Standard, heat denatured and resolved in POP4 polymer on an ABI 3100 Genetic Analyzer (all Applied Biosystems, Darmstadt, Germany). Raw data were analyzed and alleles assigned using the GENEMAPPER 4.0 software.

Paternity assignment

An integrated Bayesian analysis was implemented in MasterBayes v.2.51, available at http://cran.r-pro- ject.org/ (Hadfield et al. 2006) using a Markov Chain Monte Carlo approach to estimate paternity through a full probability model. We restricted the candidate fathers to males breeding in the same plot where a chick was born, either the same year or the previous year. We also included whether a candidate genetic father was the social male as a predictor of the probability of siring a chick. To provide additional spatial information, each individual was assigned coordinates corresponding to the nest box location. Nest box coordinates were then used within the model to specify the distance between each chick and potential sire in interaction with year (to give more weight to males breeding in the same year that the focal offspring was born). The effect of distance between chick and each potential sire depending on the breeding year were then directly estimated from the model. Analyses used 2 million iterations, burn-in of 400 000, thinning interval 2000, improper uniform priors and assumed E1 =E2 = 0.01. Given we established our nest box population in 2010, and also conducted a chick swapping experiment, paternity analysis was done slightly differently. (1) Only males that bred in the same year as a focal offspring were included. We considered potential genetic

fathers to be males breeding in the plots where the swapping experiment occured (sometimes chicks were swapped between different plots). (2) We did not include spatial information because the original nest where the chick came from was unknown. (3) Chicks could have been born in two different nest boxes (original or swap nest box), therefore as extra phenotypic information we fitted a variable that coded whether the candidate father was breeding in one of the pair of swapped nest boxes of a particular chick.

Table S1 Microsatellite loci for *Parus major*. Primer sequences include information on fluorescence labels used. *C*, primer concentration in multiplex primer mix; Ta, annealing temperature.

Locus	Accession no.	published in	designed for (original species)	Primer sequences (5' - 3')	*C* (µM)	Multiplex Mix	T_a (˚C)	Size range (bp)	number of alleles
ADCYAP1_bm	FJ464427	Steinmeyer et al (2009), supplement	for blue tit (*Cyanistes caeruleus*) on zebrafinch and chicken genome	F: VIC-GATGTGAGTAACCAGCCACT R: ATAACACAGGAGCGGTGA	0,24 µM	2A, 2B	57	154 - 167	10
POCC6	U59117	Bensch et al. (1997)	western crowned warbler (*Phylloscopus occipitalis*)	F: VIC-TCACCCTCAAAAACACACACA R: ACTTCTCTCTGAAAAGGGGAGC	0,16 µM	1A, 1B	55	172 - 216	21
Pca7	AJ279809	Dawson et al. (2000)	blue tit (*Cyanistes caeruleus*)	6FAM-TGAGCATCGTAGCCCAGCAG GGTTCAGGACACCTGCACAATG	0,17 µM	3A, 3B	58	95 - 101	4
Pca8	AJ279810	Dawson et al. (2000)	blue tit (*Cyanistes caeruleus*)	NED-ACTTCTGAAACAAAGATGAAATCA TGCCATCAGTGTCAAACCTG	1,5 µM	1A, 1B	55	185 - 239	26
Pca9	AJ279811	Dawson et al. (2000)	blue tit (*Cyanistes caeruleus*)	VIC-ACCCACTGTCCAGAGCAGGG AGGACTGCAGCAGTTTGTGGG	0,12 µM	3A, 3B	58	113 - 131	12
Mcy-H4	U82388	Double et al. (1997)	superb fairywren (*Malurus cyaneus*)	PET-ATAAGATGACTAAGGTCTCTGGTG TAGCAATTGTCTATCATGGTTTG	2,8 µM	1A, 1B	55	137 - 175	11
PmaC25	AY260526	Saladin et el. (2003)	great tit (*Parus major*)	PET-CGTCCTGCTGTTTGTATTTCTG CCATGAACCATTTTTAGGGTG	0,24 µM	1A	55	308 - 344	14
PmaGAn27	AY260532	Saladin et el. (2003)	great tit (*Parus major*)	NED-TATAAACCACAGCCACACGC CACAACCACAGAGGCATGAG	0,14 µM	2A, 2B	57	179 - 284	32
PmaTGAn33	AY260539	Saladin et el. (2003)	great tit (*Parus major*)	6FAM-TTCCCCAAGTATCCTGCATC AAACCATATCACCCAGTGCC	0,17 µM	2A, 2B	57	249 - 348	40
PmaTGAn42	AY260540	Saladin et el. (2003)	great tit (*Parus major*)	6FAM-ACTTCCACATGCCAGTTTCC TGTTAAGGCAGAGAGGTGGG	0,16 µM	1A, 1B	55	229 - 293	13
PmaTAGAn71	AY260537	Saladin et el. (2003)	great tit (*Parus major*)	NED-TCAGCCTCCAAGGAAAACAG GCATAAGCAACACCATGCAG	0,18 µM	3A, 3B	58	157 – 213	14
PmaTAGAn86	AY260538	Saladin et el. (2003)	great tit (*Parus major*)	6FAM-AAAACAAGGCCACTTAGAGCTG ACTCCTCCAGGTCACACAGG	0,16 µM	2A, 2B	57	135 - 227	37
PmaD105	AY260528	Saladin et el. (2003)	great tit (*Parus major*)	PET-CAAATCACACAGTTGCTGCC CCTGGTATAAGACTGGTCAAAACAG	0,17 µM	3A, 3B	58	378 - 438	17
PmaD130	AY260529	Saladin et el. (2003)	great tit (*Parus major*)	VIC-TGAGTGGAAAGATGCTGGC CCCTATAAAAACCGAGGCTG	0,68 µM	3A, 3B	58	374 - 462	27
ClkpolyQ	not published at GenBank	Johnsen et al (2007)	blue tit (*Cyanistes caeruleus*)	6FAM-TTTTCTCAAGGTCAGCAGCTTGT CTGTAGGAACTGTTGYGGKTGCTG	0,22 µM	3A	58	268 - 286	5
NPAS2	FJ464429	Steinmeyer et al (2009), supplement	blue tit (*Cyanistes caeruleus*)	PET-CTGTGGTAAATTTGATGATTCTGA ACACCAAGTTCTTTGCACAATG	0,38 µM	2A	57	168 - 195	9
DRD4 UTR Indel	not published at GenBank	unpublished, Jakob Mueller, Christine Baumgartner	great tit (*Parus major*)	VIC-CTGGTCTGCTGTCTTTGTTGG GGACATCTGGGAAATGAGCTT 57.9	0,22 µM	2A	57	302 - 311	4
DRD4 ID13606	not published at GenBank	unpublished, Jakob Mueller, Christine Baumgartner	great tit (*Parus major*)	PET-GCAGGACAAGTGACCCCTC 61.0 AATCAAGCCCAAGGTGAGCA	0,12 µM	2A	57	323 - 339	4
Titgata68	AY792960	Wang et al.(2005)	green backed tit (Parus monticolus)	PET-ACAGATCAGCATGGTTGCAG CATCCACAAGGGCAATCTTT	0,28 µM	1B	55	228 - 272	21
CcaTgu6	CK235244.1	Olano et al. (2010)	blue tit (*Cyanistes caeruleus*)	NED-ACAATTGCTAACAAGTGGCAAG AAGTGAAATCTKCTTGGGKC	0,32 µM	1B	55	102 - 117	6
DkiB102-ZEST	AY769673.1	Olano et al. (2010) King et al. (2005)	Kirtland's warbler (Dendroica kirtlandii)	PET-TTGCAACAGGAGGACAAGG CAGCAGCACTTCCCAATACA	0,14 µM	2B	57	194 - 272	26
CcaTgu19	DV579042.1	Olano et al. (2010)	blue tit (*Cyanistes caeruleus*)	VIC-CTGGACCATGACTGCAAGATT CAGTGGCAAAKCAGCACCT	0,14 µM	2B	57	234 - 333	27
CcaTgu27	DV947660.1	Olano et al. (2010)	blue tit (*Cyanistes caeruleus*)	6FAM-ARACAGGGCGAAGTTTCTGAR GCAGATTCATGAGATGATGAGAGA	0,58 µM	3B	58	160 - 169	4
P2/P8	AF006659-62	Griffiths et al. (1998)	chicken and zebra finch	6FAM-CTCCCAAGGA TGAGRAAYTG TCTGCATCGC TAAATCCTTT	0,44 µM	1	55	321, 373	

References

Bensch, S., Price, T. & Kohn, J. 1997. Isolation and characterization of microsatellite loci in a Phylloscopus warbler. *Molecular Ecology*, **6**, 91–92.

Dawson, D. A., Hanotte, O., Greig, C., Stewart, I. R. K. & Burke, T. 2000. Polymorphic microsatellites in the blue tit Parus caeruleus and their cross-species utility in 20 songbird families. *Molecular Ecology*, **9**, 1941–1944.

Griffiths, R. & Double, M. 1998. A DNA test to sex most birds. *Molecular Ecology*, **7**, 1071–1075.

Hadfield, J. D., Richardson, D. S. & Burke, T. 2006. Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework. *Molecular Ecology*, **15**, 3715–3730.

King, T. I. M. L., Eackles, M. S., Henderson, A. P., BOCETTI, C. I., Currie, D. & Wunderle, J. M. 2005. Microsatellite DNA markers for delineating population structure and kinship among the endangered Kirtland's warbler (Dendroica kirtlandii). *Molecular Ecology Notes*, **5**, 569–571.

Olano-Marin, J., Dawson, D., Girg, A., Hansson, B., Ljungqvist, M., Kempenaers, B. & Mueller, J. C. 2010. A genome-wide set of 106 microsatellite markers for the blue tit (Cyanistes caeruleus). *Molecular Ecology Resources*, **10**, 516–32.

Saladin, V., Bonfils, D., Binz, T. & Richner, H. 2003. Isolation and characterization of 16 microsatellite loci in the Great Tit Parus major. *Molecular Ecology Notes*, **3**, 520–522.

Sambrock, J. & Russel, D. W. 2001. Molecular Cloning: A Laboratory Manual(3rd edition).

Steinmeyer, C., Mueller, J. C. & Kempenaers, B. 2010. Search for informative polymorphisms in candidate genes: clock genes and circadian behaviour in blue tits (vol 136, pg 109, 2009). *GENETICA*, **138**, 469.

Wang, M. T., Hsu, Y. C., Yao, C. T. & Li, S. H. 2005. Isolation and characterization of 12 tetranucleotide repeat microsatellite loci from the green-backed tit (Parus monticolus). *Molecular Ecology Notes*, **5**, 439–442.

Appendix S2. Additional models for the analysis of within-pair and extra-pair paternity

Table S2. *Variance components for within-pair and extra-pair paternity from a binomial mixed effect model. Estimates presented are the mode of the posterior distribution and in parentheses the lower and upper credible interval limits (95%CI).*

Response variable	Within-pair * paternity loss	Extra-pair * paternity gain
Fixed effects	β (95% CI)	β (95% CI)
Intercept	-0.61	-1.35
	(-0.79, -0.43)	(-1.56, -1.14)
Random effects	σ^2 (95% CI)	σ^2 (95% CI)
$V_{among\text{-}males}$	0.00	0.46
	(0.00, 0.00)	(0.42, 0.54)
$V_{among\text{-}females}$	0.44	0.00
	(0.39, 0.50)	(0.00, 0.00)

*Untransformed estimates from a generalized mixed effect models with binomial error distribution,

Comparison of Poisson versus binomial models

The difference between the sources of variation of within-pair paternity loss estimated by models assuming binomial error distribution versus Poisson distributions, stem from differences in the amount of variance associated male identity effects. Our interpretation of these differences is that the probability of losing within-pair paternity is only due to female promiscuity whereas the amount of paternity loss is determined by both, female promiscuity and male ability to avoid within-pair paternity loss.

Table S3. *Relation between the different siring routes and male's annual siring success derived from a mixed effect model. We present the mode of the posterior distribution, this effects are mathematically true by definition, and therefore there is no uncertainty in the estimates.*

Response variable	Male yearly siring success
Fixed effects	β (95% CI)
Intercept	0.10
Clutch size	0.49
Within-pair paternity loss	-0.49
Extra-pair paternity gain	0.49
Random effects	σ^2 (95% CI)
$V_{among\text{-}males}$ n= 460 males	0.00
$V_{among\text{-}females}$ n=475 females	0.00
$V_{among\text{-}observations}$ n=695 breeding attempts	0.00

Table S4. *Effect of clutch size in male within-pair paternity loss derived form a mixed effect model. Estimates presented are the mode of the posterior distribution with the upper and lower limits of the credible intervals in parenthesis. Within-pair paternity loss was modeled assuming over-dispersed Poisson errors.*

Response variable	Within-pair paternity loss
Fixed effects	β (95% CI)
Intercept	-0.95 (-1.08, -0.81)
Clutch size	0.11 (0.03, 0.20)
Random effects	σ^2 (95% CI)
$V_{among\text{-}males}$ n= 460 males	0.18 (0.14, 0.19)
$V_{among\text{-}females}$ n=475 females	0.16 (0.14, 0.18)
$V_{among\text{-}observations}$ n=695 breeding attempts	0.66 (0.60, 0.74)

Table S5. *Age effects on within-pair paternity loss and extra-pair paternity gain derived from mixed effect models. Parental age is based on birth year for locally born birds or plumage characteristics at firs catching for immigrants (Svensson 1992). Immigrants first caught with adult plumage are assigned a minimal age of 2 years (following (Bouwhuis et al. 2009). Estimates presented are the mode of the posterior distribution and in parentheses the lower and upper credible interval limits (95%CI). Within-pair paternity loss and extra-pair paternity gain are modeled assuming over-dispersed Poisson error distributions.*

Response variable	Within-pair paternity loss	Extra-pair paternity gain
Fixed effects	β (95% CI)	β (95% CI)
Intercept	-1.01	-2.38
	(-1.19, -0.85)	(-2.66, -2.07)
Age	0.07	0.33
	(-0.04, 0.18)	(0.15, 0.51)
Random effects	σ^2 (95% CI)	σ^2 (95% CI)
$V_{among\text{-}males}$	0.18	0.29
	(0.16, 0.20)	(0.25, 0.32)
$V_{among\text{-}females}$	0.14	0.00
	(0.12, 0.16)	(0.00, 0.00)
$V_{among\text{-}observations}$	0.73	2.95
	(0.66, 0.81)	(2.64, 3.23)

References

Bouwhuis, S., Sheldon, B. C., Verhulst, S., & Charmantier, A. (2009). Great tits growing old: selective disappearance and the partitioning of senescence to stages within the breeding cycle. *Proceedings. Biological Sciences / The Royal Society*, *276*(1668), 2769–77. doi:10.1098/rspb.2009.0457

Svensson, L. (1992). *Identification guide to European passerines*. L. Svensson.

Appendix S3. Fitting multivariate models in lme4.

The package lme4 is not conventionally used to parameterize multivariate mixed effect models, but it allows fitting random regression models with a nested structure (Bates et al. 2014). Re-parameterizing a random regression model is possible to quantify covariances between different traits at different levels, using this package's restricted maximum likelihood algorithm. Random regression models allow for the estimation of correlations (hence also covariances) between intercepts and slopes. For example, the correlation between an individual's intercept deviation from the population mean intercept with the individual's slope deviation from the population mean slope. If the environmental gradient has only two levels it is possible to estimate an intercept for each level and estimate the correlation across levels, instead of estimating the correlation between the intercept and the slope of the environmental gradient. This is mathematically equivalent to multivariate mixed effect model with two traits (response variables). To estimate the within-individual correlation from a Poisson model, it is simply necessary to add a separate over-dispersion parameter for each of the two environments (traits) and estimate their correlation. The advantage of utilizing this approach is that it enabled us to perform mixed-effect model estimating the point estimates with restricted maximum likelihood framework with the possibility to estimate posterior distributions (Gelman & Hill 2007).

We performed a simulation study to ensure that this approach was not producing biased estimates. We simulated data sets with the same level of replication as in our study. The datasets consisted of repeated measures of two correlated response variables for 454 individuals, where all individuals were measured in only one year, 38% in both of two years, 13% in all of three years and 9% in all of four years. The response variables were correlated at the among- and within-individual level. We also modeled an over-dispersed error distribution (as in our dataset). We simulated 100 datasets with different correlation patterns at the among- and within-individual levels. We varied independently the among-individual correlations and the within-individual correlations across a range of values (-0.7, -0.5, -0.3, -0.1, 0.1, 0.3, 0.5, 0.7). To each of these data sets we then fitted a model to estimate the among- and within-individual correlations, and estimated the bias in parameter estimates for each combination of simulated parameters (among- and within-individual correlations). We used the package MASS to simulate the data sets assuming variances of 0.3 for the

among-individual variance of both traits and 0.7 for the within individual variance of both traits. To assess model performance, we estimated bias, imprecision, and the probability that the estimated value was of different sign. Bias was quantified as the absolute mean difference of the estimated parameters of 50 simulations with the simulated parameter. We also measured the impression of the models as the standard deviation of the 100 estimates for each combination of parameters (within- and among-individual correlations). We also assessed whether the probability that the estimated correlations differed in sign compared to the correlation set to simulate the data.

In summary our statistical models estimate correlations with low bias (the maximum bias was 0.15; Fig S1). Regarding imprecision, weak correlations at the among-male level are more imprecise and there is a higher chance that the estimated correlation will be of opposite sign than the simulated correlation (Fig S2, S3). The within-individual correlations are very precise and the probability that the estimated correlation is of a different sign is low even for very weak correlations (Fig S2, S3). Our simulations show that are approach returned unbiased estimates with an accuracy or precision not much different if we will have used another software to fit the models (for an example see; Dingemanse & Dochtermann 2013)

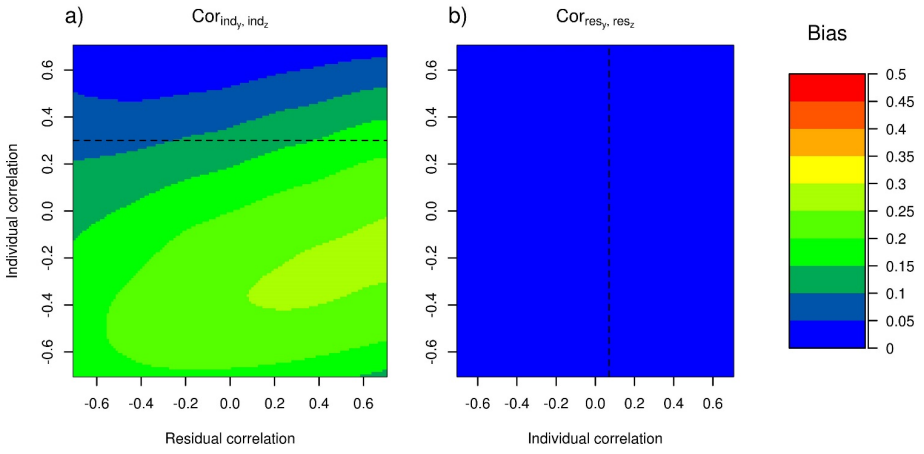

Figure S1. *Bias of multivariate mixed effect model estimates of (a) among- and (b) within-individual correlations as a function of different magnitudes of among- and within-individual correlations. Models were applied to simulated data sets within the parameter space of among- and within-individual correlations, between -0.7 and 0.7. The different colours depict areas between isoclines of similar levels of inaccuracy; isoclines were determined by bilinear interpolation between the sampled integer values of the different correlation magnitudes. Dotted line represents the estimated correlations.*

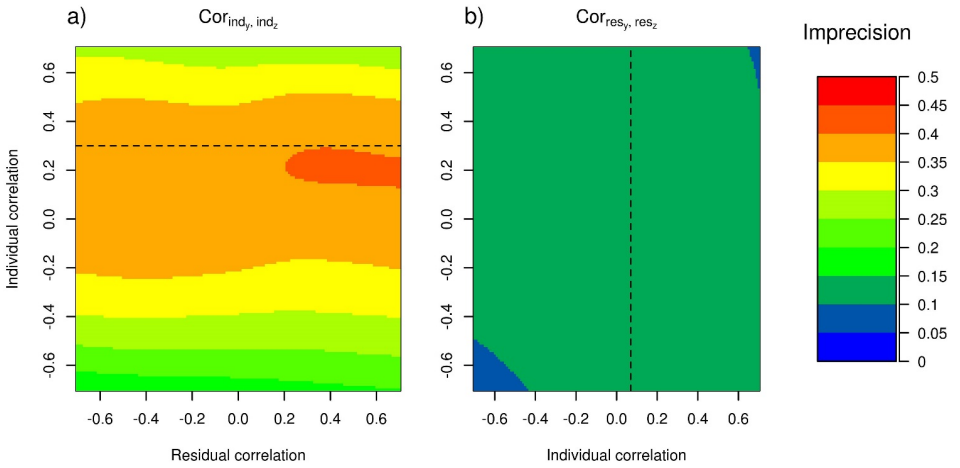

Figure S2. *Imprecision of multivariate mixed effect model estimates of (a) among- and (b) within-individual correlations as a function of different magnitudes of among- and within-individual correlations. Models were applied to simulated data sets within the parameter space of among- and within-individual correlations, between -0.7 and 0.7. The different colours depict areas between isoclines of similar levels of inaccuracy; isoclines were determined by bilinear interpolation between the sampled integer values of the different correlation magnitudes. Dotted line represents the estimated correlation in our empirical study.*

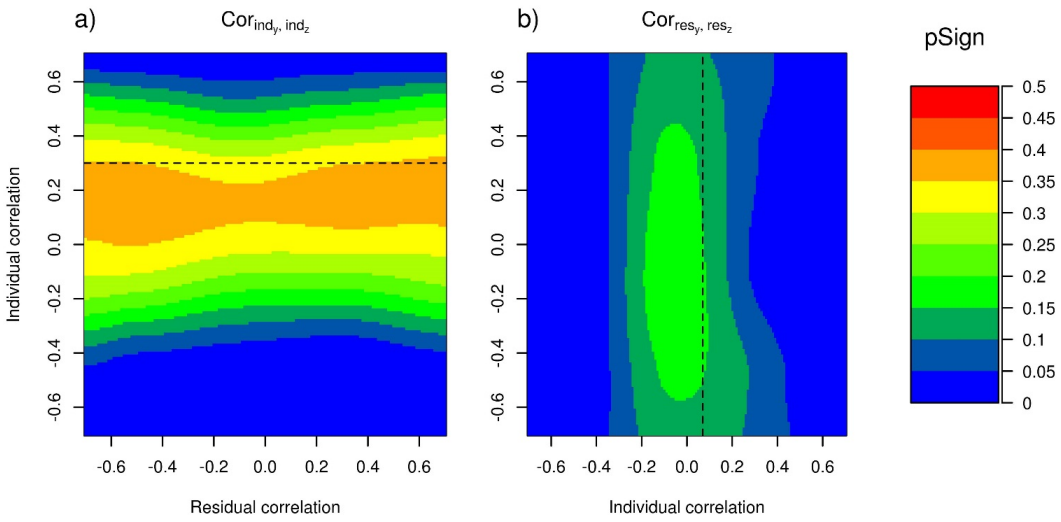

Figure S3. *Probability that the estimated correlation is of different sign from the simulated correlation of multivariate mixed effect model estimates of (a) among- and (b) within-individual correlations as a function of different magnitudes of among- and within-individual correlations. Models were applied to simulated datasets within the parameter space of among- and within-individual correlations, between -0.7 and 0.7. The different colours depict areas between isoclines of similar levels of inaccuracy; isoclines were determined by bilinear interpolation between the sampled integer values of the different correlation magnitudes. Dotted line represents the estimated correlation in our empirical study.*

References

Bates, D., Mächler, M. & Bolker, B. 2014. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*,

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

Gelman, A. & Hill, J. 2007. *Data analysis using regression and multilevel/hierarchical models.* New York: Cambridge University Press.

Chapter 5

Timing of extra-pair fertilizations: within-pair fertilization trade-offs or pair synchrony spill-overs

Yimen G. Araya-Ajoy, Niels J. Dingemanse, Bart Kempenaers

Manuscript

Abstract

Extra-pair copulations are expected to increase male reproductive success unless they constrain a male's ability to secure within-pair fertilizations. Therefore, natural selection should favor males that engage in extra-pair copulations when it does not interfere with their ability to secure within-pair fertilizations, for instance, when their social female is not fertile. On the other hand, the timing of extra-pair fertilizations (EPF) may be constrained by a male's need to synchronize his reproductive cycle with his social female's. For example, to ensure within-pair fertilizations, a male's sperm production and willingness to copulate should peak during the fertile period of his social female. This will lead to higher male willingness and/or success in siring extra-pair offspring when his social female is fertile, due to a spill-over effect of his within-pair reproductive behavior. We studied the timing of extra-pair fertilizations in male great tits (*Parus major*) to determine if males time their extra-pair fertilizations to avoid a trade-off with securing within-pair fertilizations (trade-off avoidance hypothesis), or if males within-pair fertilization behavior spills over to his extra-pair fertilization behavior, causing extra- and within-pair fertilizations to occur around the same time (pair synchrony spill-over hypothesis). It is known that extra-pair reproduction is also determined by the availability of fertile females in the vicinity (extra-pair fertilization opportunity). As expected, we found that extra-pair fertilization opportunity determined the probability of fertilization success. However, when correcting for variation caused by differences in opportunity, a male's extra-pair fertilization success was highest when his social female was fertile. This result supports the idea that a male's within-pair fertilization behavior spills over to his extra-pair fertilization behavior, causing most extra-pair sirings to be obtained when his social female is fertile. Given that the exact fertile period length of a female great tit is not known, we studied the effect of different fertile period lengths on our interpretation of the observed temporal patterns of extra-pair fertilizations. Irrespective of the assumed fertile period length, males were more likely to gain extra-pair paternity when their social female was fertile. Moreover, our results show that if a female's fertile window is very short, males exploit their extra-pair fertilization opportunities disproportionately more when their social female is fertile compared to when she is not. In conclusion, the results of this study support the pair synchrony spill-over hypothesis.

Introduction

For a socially monogamous male, the costs and benefits of engaging in fertilization related activities should vary as a function of the fertile cycle of his social female. The costs of keeping his reproductive machinery at an optimum results in seasonal variation in reproductive physiology and behavior, and the efficacy of fertilization attempts across the fertile cycle of his female shapes this seasonal variation. Therefore, to increase copulation and fertilization success, a male's physiology and behavior should be finely tuned to the fertile cycle of his social female. For example, a male's copulation rate and advertisement displays peak when his social female is most fertile (Birkhead et al. 1987; Mace 1987). Also, a male's aggressive reaction towards a territorial intruder is more intense when his social female is fertile compare to when she is not (Araya-Ajoy & Dingemanse 2014), presumably to avoid being cuckolded. Why males tune their fertilization-related behaviors to the fertile cycle of their social female is clear, but how it affects males' extra-pair fertilization behaviors is still unclear. Male reproductive physiology and fertilization related behaviors are strongly affected by his social female's fertile cycle, therefore it is expected that male extra-pair fertilization behavior is also affected. This may result in similar or opposite temporal patterns of investment in within-pair versus extra-pair reproduction, depending on the existence of conflicts between these alternative routes to fertilization success (as we detail in chapter four). There are no empirical studies about the relation between a male's extra-pair siring success and his social female's fertile cycle. Therefore, in this study we aimed to address the question of how males' extra-pair siring success is affected by the fertility cycle of their social female.

One of the most important determinants of male extra-pair fertilization success in birds is the availability of fertile females in the vicinity (Westneat & Stewart 2003; Mayer & Pasinelli 2013). Thus, the timing of a male's extra-pair fertilization success in relation to his social female's fertile cycle will be determined by the amount of fertile females available. For example, if a male has a social female that is fertile earlier than the majority of females in the population, he will have most of his extra-pair fertilization opportunities after the fertile period of his social female. On the other hand, a male with a social female that is fertile late in the breeding season will have most of his extra-pair opportunities before his social female is fertile. Therefore, the timing for a male to engage in extra-pair reproduction in relation to his female's fertility stage is partly determined by the patterns of opportunity arising from when his

social female is fertile compared to other females in the population. It is possible that after accounting for the available opportunities in the different fertile stages of his social mate, a male's extra-pair paternity success will not be different across the different fertile stages of his social female (Figure 1a). We set this pattern as our null hypothesis, against which we tested two alternative hypotheses: the “trade-off avoidance” hypothesis and the “pair synchrony spill-over” hypothesis (both detailed below).

In the majority of socially monogamous species, most of male reproductive success stems from within-pair reproduction (Webster et al. 1995; Schlicht & Kempenaers 2013; Lebigre et al. 2013). Therefore, males should invest in extra-pair fertilizations when it does not conflict with securing within-pair fertilizations. For instance, males should invest in securing within-pair paternity when their social female is fertile and invest in extra-pair fertilizations when she is not fertile (i.e. when there is no chance of losing within-pair paternity). Therefore, under the “trade-off avoidance” hypothesis, we predict that (after correcting for variation in opportunity) males will sire fewer extra-pair offspring during their social female's fertile period (Figure 1b). Several mechanisms could be underlying this pattern: either males are less likely to engage in extra-pair copulations during the fertile period of their social female or extra-pair copulations are less likely to fertilize an egg. For instance, a male's investment in mate guarding during the fertile period of his female may constrain his investment in searching for extra-pair copulations (Westneat et al. 1990; Kokko 2005) and this may lead to fewer extra-pair sirings during this period. Another possibility is that sperm depletion due to frequent copulations with his own mate (Westneat et al. 1990; Birkhead 1991) could cause extra-pair copulations to be less likely to fertilize an extra-pair egg (due to lower sperm count) when his social female is fertile.

It is also possible that males are not able to optimally time their extra-pair fertilizations to avoid potential trade-offs between within-pair and extra-pair fertilizations. For instance, in order to secure within-pair fertilization success, a male's willingness to copulate and sperm production should peak during the fertile period of his social female; this will increase his within-pair reproductive success, but could also spill over to his extra-pair siring success. This pattern will result from a male's need to synchronize his breeding physiology and behavior with his social female to ensure within-pair fertilization success (Fusani 2008).Therefore,

under the “pair synchrony spill-over” hypothesis we predict that males will sire more extra-pair offspring when his social female is fertile (Figure 1c). Two mechanisms could mediate this process: either males are more likely to engage in extra-pair copulations during the fertile period of their social female, or extra-pair copulations are more successful at this time. For example, male behaviors directed to increase within-pair fertilizations could incidentally increase his chances for extra-pair fertilizations (Figure 1d). It has been shown that the advertisement display in male birds peaks when their mates are fertile in order to increase their within-pair fertilization success (Mace 1987; Halfwerk et al. 2011). This could attract other females as well, and therefore also increase males' extra-pair fertilization success (Kempenaers et al. 2010). Another possibility is that a male's reproductive capacity is highest when his social female is fertile (e.g. sperm quality, sperm number), and hence extra-pair copulations would be more likely to result in successful extra-pair fertilizations.

We studied the timing of male great tits' (*Parus major*) extra-pair reproduction in relation to the fertility status of their social female. We wanted to test two mutually exclusive hypotheses: males engage in extra-pair copulations when it does not trade off with securing within-pair fertilizations (trade-off avoidance hypothesis) or male timing of extra-pair fertilizations is driven by a spill-over effect of their within-pair fertilization behavior (pair synchrony spill-over hypothesis). To address these hypotheses, we quantified a male's opportunities and realized extra-pair fertilization success before, during, and after the fertile period of his social female. First, we showed how a male’s opportunity to sire extra-pair offspring changes depending on the seasonal start of the fertile period of his social mate. Then, we assessed whether realized extra-pair fertilizations simply reflected opportunity (H_0, Fig. 1a). We expected that an increase in opportunity would increase the chances of extra-pair fertilization success in all three periods (before, during, and after his social female is fertile). We then determined when male's extra-pair fertilization success was highest in relation to his social female's fertility status, after correcting for variation in opportunity. If males gain less extra-pair paternity (relative to opportunity) during the period when their social female is fertile, the trade-off hypothesis would be supported (H_{A1}, Figure 1b). On the other hand, the data will support the “pair synchrony spill-over” hypothesis if males gain more extra-pair paternity during the fertile period of their social female (H_{A2}, Figure 1c). To be able to perform these analyses, it is necessary to know the length of the fertile period of female great tits. Precise information about the start of the fertile period is unfortunately not available,

though behavioral patterns suggest that the days before the first egg is laid are of key importance (Birkhead et al. 1987; Michl et al. 2002; Forstmeier et al. 2011). To determine the effects of different possible lengths of female fertile period on the support for each alternative hypothesis, we studied how different fertile period lengths would affect our interpretation of the observed temporal patterns of extra-pair fertilizations.

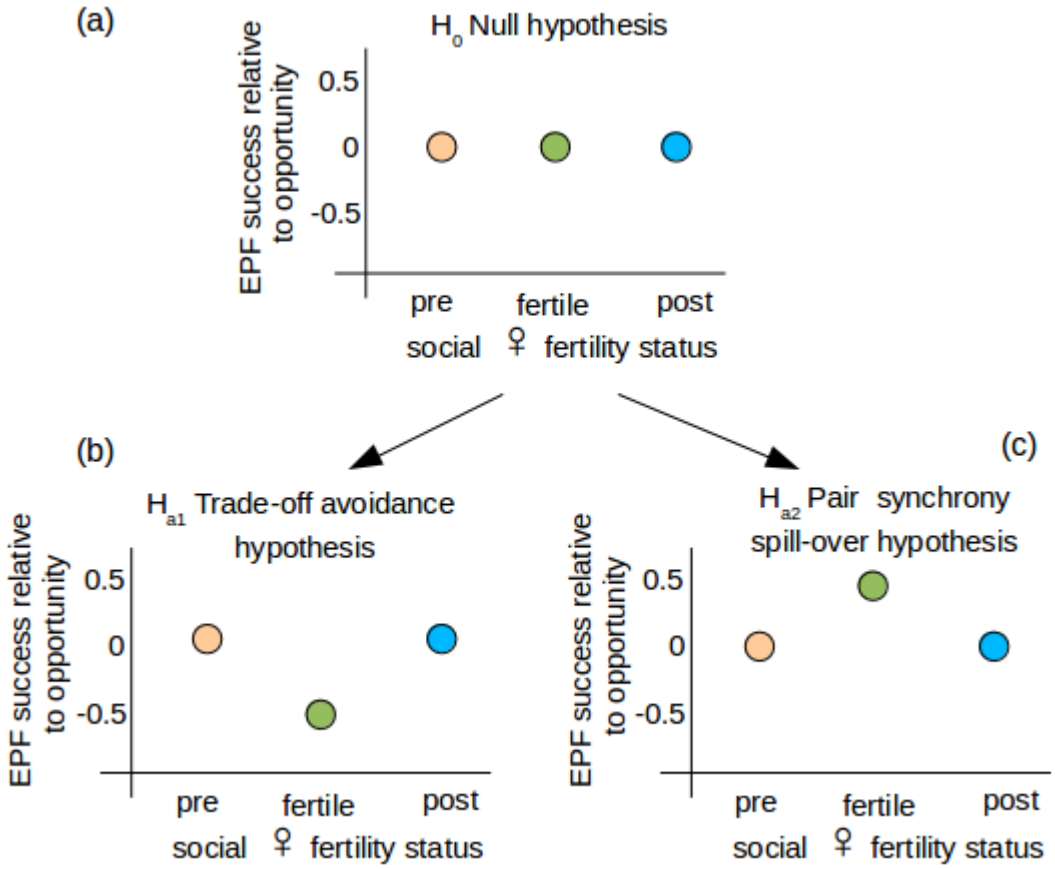

Figure 1. *Schematic representation of the hypotheses and predictions regarding males' timing of extra-pair fertilization success in relation to their social female's fertile status. (a) Represents the null hypothesis, which predicts that extra-pair paternity is affected by the available opportunity. After correcting for variation in opportunity, a male's probability of extra-pair fertilization success is the same across the different fertile stages of his social female. (b) The trade-off avoidance hypothesis; states that trade-offs between within-pair and extra-pair fertilizations cause individuals to have a lower extra-pair fertilization success, relative to the available opportunity, when their social female is fertile. (c) The "pair synchrony spill-over" hypothesis predicts that, after accounting for variation in opportunity, male-fertilization success is higher when his social female is fertile.*

Methods

Population and study site

We studied 12 nest box plots of great tits in Southern Germany during 4 years (2010-2013), in an area of approximately 120 ha (Bavarian Landkreis Starnberg; 47º 58' N, 11º 14' E). Each plot consisted of a regular grid of 50 boxes, with 50 meters between adjacent boxes. From April onwards, boxes were checked twice a week to determine lay date (back-calculated assuming that one egg was laid per day) and final clutch size. When the nestlings were 6 days old, they were marked with an aluminum ring, blood samples were taken, and any unhatched egg was collected. Parents were caught with a spring trap in the nest box the next day, measured, bled, and marked with a unique combination of rings if caught for the first time.

Genotyping and assignment of parentage

Because we were interested in male siring success, we tried to genotype most of the produced offspring (cf. successfully hatched nestlings, unhatched eggs, and nestlings deceased prior to blood sampling) in the first clutches of all males in our population (nests starting within 30 days after the first egg of the focal year in each plot was found). We recorded 6722 eggs in our population and managed to genotype 53447 offspring. We excluded all breeding attempts where the maternity was uncertain (social mother not caught) and used a 90% confidence as a cut-off to take a paternity assignment to further analyses (see chapter four for further details). Genetic parentage was assigned using genetic and spatial information incorporated in Bayesian full probability models (R package MasterBayes; Hadfield et al. 2006). We assigned paternity to 4018 offspring (75 % out of 5,347 genotyped offspring) from 454 males and 466 female parents in 668 breeding attempts (141 breeding attempts in 2010, 163 in 2011, 223 in 2012, and 158 in 2013).

Estimating opportunity in relation to social female breeding stage

We estimated a male's opportunity to gain extra-pair paternity during three fertility stages of his social female: before, during, and after she is fertile (Figure 2). We measured opportunity as “female fertile days”, which were defined as the sum of days when all the females in the vicinity of a focal male were fertile during a particular period. For example, if there were four possible extra-pair females for a particular male, fertile during three days when his social female was also fertile, he will be assigned an opportunity of 12 fertile days during the fertile

period of his social female. Females available for extra-pair reproduction for each male's breeding attempt were restricted spatially to females breeding in the same plot and year as the focal male. Given the spatial distribution of our plots it is very unlikely that a male will sire any extra-pair offspring in another plot (General Introduction; Figure 4). The two plots that are closest to each other are about 1.2 km apart, and we found no evidence that a male breeding in one plot would have sired an extra-pair offspring in the other plot. The opportunity for a particular male to gain extra-pair paternity before the fertile period of his social female was calculated as the sum of all the female fertile days before his social female was fertile. A male's opportunity to gain extra-pair paternity during the fertile period of his social female was calculated as the amount of female fertile days that overlapped with the fertile period of his social female. A male's opportunity after his social female's fertile period was defined as the sum of female fertile days after the fertile period of his social female was over.

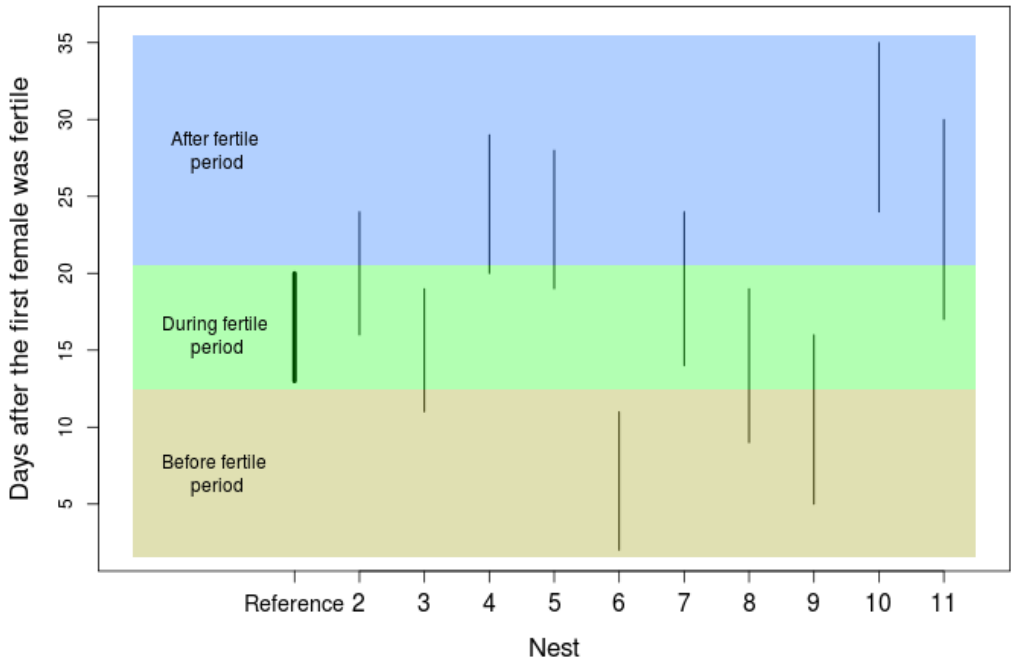

***Figure 2**. Graphical representation of how male's extra-pair fertilization opportunity was calculated in relation to his social female's fertile stage. Each vertical line represents a nest and the length of the line represents the fertile period length. Vertical lines represent the fertile period of a particular nest and the bold line is a randomly chosen nest for reference. We depict one plot in a particular year in reference to the fertile stage of the female from the reference nest (thick vertical line). Lines or part of lines overlapping the shaded blue area represent the opportunity to gain extra-pair fertilizations of the male from the reference nest after the fertile period of his social female. Lines overlapping the shaded green area represent the opportunity to gain extra-pair paternity during the fertile period of his social female and lines overlapping the brown shaded area represent the opportunity to gain paternity before his social female is fertile.*

Estimating extra-pair paternity timing in relation to social female fertility stage

We estimated a male's extra-pair fertilization success before, during, and after the fertile period of his social female. A male's fertilization success before the fertile period of his social female was defined as the extra-pair chicks sired with females that were fertile before the onset of his social female's fertile period. A male's fertilization success during the fertile period of his social female was defined as the extra-pair chicks sired with females whose fertile period overlapped with the one of his social female. A male's fertilization success after the fertile period of his social female was defined as the extra-pair chicks he sired with females that were fertile after his social female. Sometimes the fertile period of an extra-pair female overlapped with different fertile stages of the social female of a particular male, and in these cases the extra-pair offspring was assigned to the fertile stage that overlapped the most with the fertile period of the extra-pair female. For example, a male sired an extra-pair offspring with an extra-pair female whose fertile period overlapped by three days with his social female's, but the extra-pair female was still fertile for two more days after the social female's fertile period ended. In this case, the extra-pair offspring will be assigned as sired during the fertile period of the social female.

The fertile period length of female great tits

To understand the effects of different fertile period lengths on the support for each alternative hypothesis, we explored how different fertile period lengths could lead to a different interpretation of the observed temporal patterns of extra-pair fertilizations. Behavioral patterns suggest that the days before the first egg is laid are of key importance to male within-pair and extra-pair fertilization success (Birkhead et al. 1993; Michl et al. 2002; Forstmeier et al. 2011). For instance, female blue tits leave their nests earlier as egg laying approaches (Schlicht et al. 2014), and the dawn chorus of male great tits peaks a few days before his social female lays the first egg (Halfwerk et al. 2011). Moreover, it is known that extra-pair offspring are more likely in the first laid eggs (Krist et al. 2005; Magrath et al. 2009; Schlicht et al. 2012), also suggesting that the days before the first egg is laid are important in extra-pair fertilizations. It has also been shown for several species that copulation rate drops dramatically after the first egg is laid (Birkhead et al. 1987), and therefore we considered different fertile period lengths beginning a few days before the first egg was laid and ending the day before the first egg was laid. The shortest period we considered began 3 days before the first egg was laid and ended the day before the onset of egg laying (fertile period length of

3 days). We then extended this period to five and seven days before the start of egg laying until the day before the first egg was laid (fertile period lengths were 5 and 7 days respectively). We then considered the possibility that extra-fertilizations can occur after the first egg was laid. Fertilization usually occurs within one hour of ovulation, which in turn usually occurs about 24 hours before the egg is laid (Birkhead et al. 1993). Given that great tits lay their eggs early in the morning, the latest possible day that a female great tit is fertile should be the day before the last egg is laid. Therefore, we consider fertile periods from three, five, and seven days before the laying of the first egg to one day before the last egg was laid. It is worth noting that assuming that the last fertile day of a female great tit is the day before the last egg is laid introduces among-female variation in the length of the fertile period due to variation in clutch size.

Statistical analyses

Effects of lay date on opportunity and realized extra-pair fertilizations

Our null model hinges on the assumption that the available opportunity is an important factor determining the temporal patterns of male extra-pair fertilizations. A male's opportunity to engage in extra-pair fertilizations in the different fertility stages of his mate is determined by when his social female is fertile compared to the rest of the females in the population. Consequently, realized patterns of a male's extra-pair paternity in relation to his social female's fertile stage should be partly determined by when his social female is fertile compared to the other females in the population. We used generalized linear mixed effect models to study these relationships. A male's opportunity and realized extra-pair paternity were modeled as a function of his social female's fertility stage (3 level factor), lay date (as a proxy of the time a female is fertile; fitted as a continuous variable), and their interaction. Lay date was mean centered by the mean lay date of each plot in each year, and its linear and quadratic effect were fitted as predictors. Random intercepts for breeding attempt (668 levels) were fitted because each breeding attempt had 3 “observations” of opportunity and realized extra-pair paternity (before, during, and after his social female's fertile period). This resulted in a total of 2,004 observations of male's fertilization opportunity and success across the different fertile stages of the social female. Random intercepts were also included for male identity (454 levels) because we had repeated measures per male across the four years.

Finally, we also included a random intercept for each combination of plot and year (plot-year, 48 levels) to fully account for any (interacting) spatial and temporal effects. Opportunity to gain extra-pair paternity was modeled assuming normal errors and realized extra-pair paternity was modeled as a binary variable (no extra-pair fertilizations = 0, more than one extra-pair fertilization =1). We modeled this process as binary because only on very few occasions did a male sire more than one chick in any particular fertile stage of his social female.

Trade-off vs. pair synchrony spill-over hypothesis

To test our hypotheses, we modeled a male's probability of extra-pair fertilization success as a function of his social female's fertile status and the amount of opportunity available (Model 1). We modeled extra-pair paternity gain as a response variable with binomial errors (with a denominator of 1), and included as fixed covariates: female fertility status (3 level factor), the male's opportunity to gain extra-pair paternity (continuous variable), and their interaction. The parameter estimates associated with effects of a social female's fertile status can be interpreted as a male's probability of fertilizing an extra-pair egg given an equal opportunity in the different fertility stages of his social female. Because extra-pair fertilization opportunity was mean centered, effect sizes are calculated at the mean opportunity of the population. The effect of opportunity on realized extra-pair paternity could be interpreted as the rate at which individuals are able to exploit the available opportunity. Finally, the interaction between female fertility status and extra-pair opportunity allows for testing whether males exploit opportunity differently in the different fertile stages of their social female. We also included random intercepts for breeding attempt identity (668 levels), male identity (454 levels), and plot-year combination (48 levels). The total sample size was 2,004 observations of male fertilization success across the different fertile stages of social females. We did not know the exact fertile period length of female great tits, and therefore we studied the effect of different fertile period lengths (detailed above) on the outcomes of our models.

General modeling procedures

Data manipulation, statistical models, and output graphics were performed in R statistical environment (R Core Team 2014). All models were fitted with the package lme4 (Bates et al. 2014) and posterior distributions were simulated using the sim function of the package arm (Gelman & Hill 2007). We present means and 95% credible intervals as descriptions of the

parameter estimates. Statistical significance was evaluated using the 95% credible intervals, and we considered a parameter estimate significant when its 95 % credible intervals did not overlap with zero.

Results

Effects of lay date on the opportunity and realized EPF success

As expected, we found that the onset of a female's fertile period (calculated as lay date) affected her mate's opportunity and realized extra-pair siring success during her different fertility stages (Figure 3, Table S1). For males with females that started laying early in the season, opportunity and realized extra-pair siring success was highest after their female was fertile and lowest before she was fertile. On the contrary, males with females that started laying late in the season had the highest opportunity and realized extra-pair siring success before their female was fertile and it was lowest after her fertile period. Finally, males with females that started laying around the average population lay date had both the highest opportunity and extra-pair siring success during the fertile period of their social female (Figure 3).

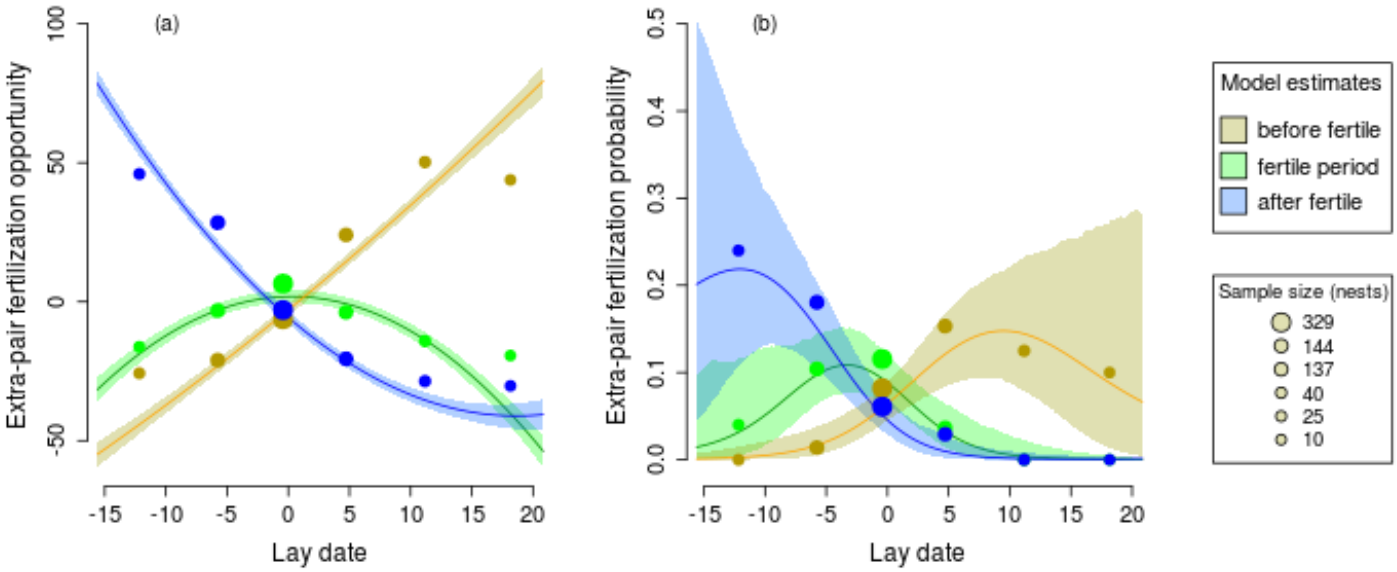

Figure 3. Effects of lay date on (a) males' opportunity for extra-pair fertilizations and (b) probability of extra-pair fertilization success in relation to different fertile stages of their social female. Lines represent model predictions and shaded areas represent the 95% credible intervals. Filled circles represent the mean values of different laying date intervals. Circle size reflects the sample size in each of the laying date intervals. Brown depicts values before fertile period of the female, green during the fertile period, and blue after the fertile period. Lay date is mean centered, and therefore negative

values represent nests that started before population's average lay date and positive values after.

Trade-off avoidance versus pair synchrony spill-over hypothesis

Within the same statistical model, we addressed two parts of our study; we tested whether the available opportunity to engage in extra-pair copulations affected the probability of gaining extra-pair paternity (validity of our null model) and we also studied when males gained more extra-pair paternity in relation to the fertile cycle of their social female. This allowed us to test our two alternative hypotheses (“pair synchrony spill-over” versus the “trade-off avoidance” hypotheses) against our null model. We explored the effect of fertile period length on the results of this model by performing these analyses assuming different fertile period lengths. We found, as expected, that an increase in opportunity increases the chance of extra-pair fertilization success (Table S2, Figure 4), and this result was consistent across the different fertile period lengths. We also found that, after correcting for differences in opportunity, individuals were more successful at siring extra-pair chicks during the fertile period of their social female, supporting the “pair synchrony spill over” hypothesis. This result was also independent of the assumed length of female fertile period (Figure 4a). Moreover, when we explored in detail the effect of the different fertile period lengths on how individuals exploit the opportunity to obtain extra-pair fertilizations, we found that if a female’s fertile period is very short (5 days or less), males exploit extra-pair reproduction opportunities more during this time as compared to other stages (Figure 4b).

Discussion

We studied the timing of extra-pair fertilizations in male great tits with relation to the fertile cycle of their social female. We aimed to determine if males engage in extra-pair reproduction when there is no trade-off with securing within-pair fertilizations (“trade-off avoidance” hypothesis) or if the temporal patterns are driven by a spill-over effect of a male's need to synchronize his breeding activities with his social female (“pair synchrony spill-over” hypothesis). For this we quantified a male's opportunity and extra-pair fertilizations success before, during, and after his social female's fertile period. Our null model was based on the assumption that the available opportunity determined extra-pair fertilization success and the support of our alternative hypotheses was based on males having a higher extra-pair fertilization success relative to the available opportunity (Figure 1).

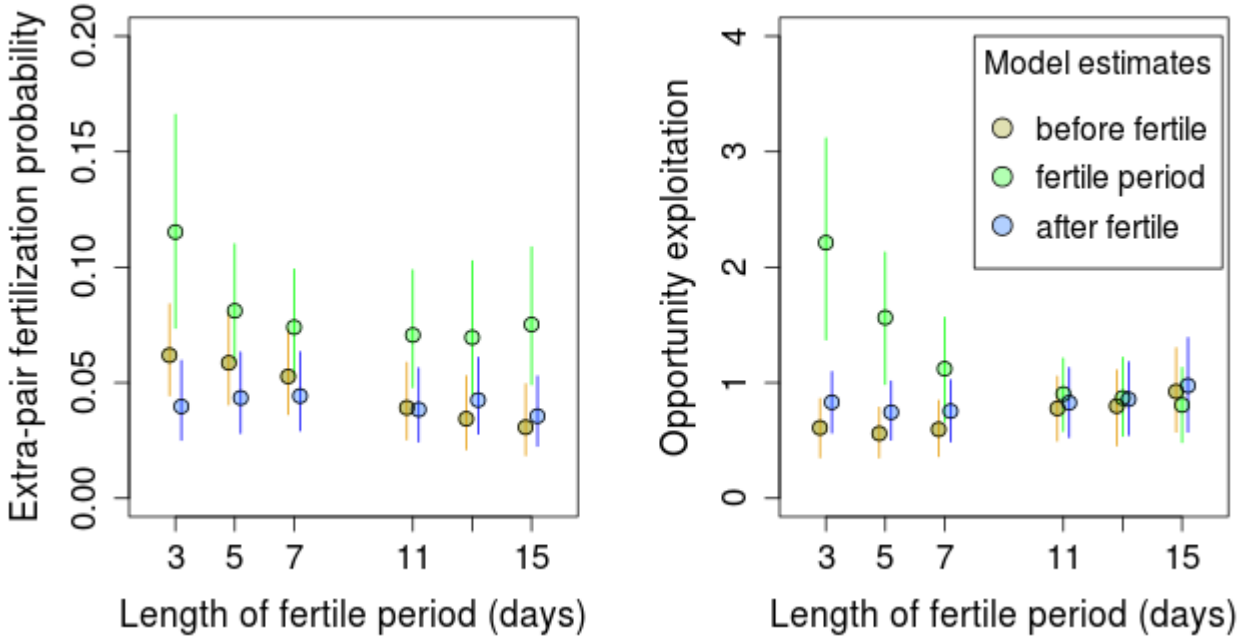

Figure 4. *(a) Extra-pair fertilization probability (y-axis) during the different social female's fertile stages (color scheme) as a function of fertile period length (x-axis). (b) Effect of opportunity on the probability of extra-pair fertilization success (y-axis) during different fertile stages of the social female (color scheme) as a function of fertile period length (x-axis). Circles refer to point estimates and bars to the 95% credible intervals. Brown depicts estimates before the fertile stage, blue after the fertile stage, and green during the fertile stage. Fertile periods from 11 to 15 days are average lengths of the population, because these fertile periods include among-female variation in clutch size (see methods). Fertile period from 3 to 7 days do not include variation associated to clutch size differences, and therefore all the females in the population have the same value.*

As expected, a male's opportunity to engage in extra-pair fertilizations during the different fertile stages of his social female depends on when his social female is fertile relative to other females in the population (Figure 3a). The realized patterns of extra-pair fertilization success followed the same pattern as extra-pair fertilization opportunity (Figure 3b). Males that had nests early in the season mostly sired extra-pair offspring with females that were fertile after their social female, while males breeding late in the season mostly sired extra-pair young with females that were fertile before their social female. Consequently, when modeling the effects of opportunity on extra-pair reproduction, we found that individuals gained more extra-pair paternity when there was more opportunity. We found that, after correcting for variation in opportunity, males' extra-pair fertilization success is higher when their social female is fertile,

supporting the pair synchrony hypothesis (Figure 4).

Length of the fertile period

When we studied how variation in the assumed fertile period length of female great tits affected the interpretation of the observed temporal patterns of extra-pair fertilization success, we revealed a very interesting pattern. Independent of the assumed fertile period length, male extra-pair siring success was always higher during the fertile period of his social female (Figure 4a). Interestingly, in a scenario where female fertile period is shorter than five days, males would also exploit extra-pair reproduction opportunity disproportionately more during the fertile period of his female (Figure 4b).

We considered six different lengths of female fertile period: three of the six assumed that the last possible fertile day was the day before the first egg was laid (fertile periods of three, five, and seven days). These fertile periods assumed that all female great tits had the same fertile period length and this assumptions was based on the observation that copulation rate decreases dramatically after the first egg is laid in some species (Birkhead et al. 1987). The other three fertile periods assumed that the last day of the fertile period was one day before the last egg was laid. This assumption introduces among-female variation in the length of the fertile period due to variation in clutch size. This situation may be more plausible, because even if copulation rate decreases after the first egg is laid, the actual fertilization of the eggs takes place about 24 hours before each egg is laid (Birkhead et al. 1993). Therefore, even if males copulate very little after the first egg is laid, females with bigger clutches will still have more possible days in which stored sperm can fertilize an egg. It has also been shown in a closely related species that copulation rate does not decrease after egg laying (Kempenaers et al. 1992) and sperm numbers found in fertilized eggs do not decrease with laying order (Johnsen et al. 2011), suggesting that the days after the first egg is laid are also important in fertilization. It is also possible that the fertile period of female great tits is like a probability distribution, and in some species it has been modeled in this way based on behavioral data (Forstmeier et al. 2011). Detailed behavioral and physiological information about great tits is needed in order to improve our knowledge of the fertile period of females. Despite the lack of knowledge about the length of a female great tit fertile period, our results support the pair synchrony hypothesis in all the different scenarios we considered.

Pair synchrony hypothesis

Most of the variation in male fertilization success stems from within-pair reproduction (Webster et al. 1995; Schlicht & Kempenaers 2013; Lebigre et al. 2013, Chapter four in this thesis). Therefore, selection should have favored male behaviors that enhance within-pair fertilization success. Male behaviors directed to increase within-pair fertilization could spill over to his extra-pair fertilization behavior, increasing the chances of male extra-pair fertilization success when his social female is fertile. During the dawn chorus, for instance, the singing output of males of several species peaks a couple of days before the onset of their social female's egg laying phase (Mace 1987; Halfwerk et al. 2011). Incidentally this increased advertisement display will be more likely to attract other females and increase their extra-pair reproductive success (Kempenaers et al. 2010). Proximately, the reproductive behavior of a female can be a cue to her social male's physiological machinery to trigger behaviors directed to increase within-pair reproductive success (Fusani 2008). Similarly, female social cues could affect male sperm production, resulting in bigger and/or better quality ejaculates during the fertile period of their social female, which will increase the success of within-pair copulations but also the success of extra-pair copulations. Our study suggests that female fertility cues trigger male reproductive physiology and behavior to enhance within-pair fertilization success, but incidentally they also affect his extra-pair reproductive behavior.

Trade-offs between within-pair and extra-pair fertilizations

It has been suggested that trade-offs between within-pair and extra-pair reproduction could cause these two activities to be temporally incompatible. For instance, in some species it has been suggested that a male's increased within-pair copulation rate could result in sperm depletion, constraining the sperm available for extra-pair copulations (Birkhead 1991). Our results suggest the contrary for male great tits, because males seem to be more successful in their extra-pair copulations at the same time that their copulation rate with their social female should be highest. It has also been suggested that a male's investment in mate guarding could constrain his ability to look for extra-pair copulations (Jennions & Petrie 2000; Kokko 2005; Westneat et al. 2009). Males may be able to resolve this trade-off by looking for extra-pair copulations when they do not have to invest in mate guarding. Our results suggest that males are not engaging more in extra-pair copulations when their females is not fertile,

but we cannot eliminate the possibility that they do engage more in extra-pair copulations but unsuccessfully.

Conclusions

We have shown that a male's extra-pair siring success is highest when his social female is also fertile. We argue that this could be the result of a spill-over effect of male's within-pair reproductive behavior to his extra-pair reproductive behavior. We were able to reveal this pattern because we accounted for variation in the available opportunity, and highlighted the importance of acknowledging that different processes can affect males' extra-pair reproduction.

Acknowledgements

We thank all the members of Evolutionary Ecology of Variation group, all field assistants and students for help in data collection and discussion, and members of the Department of Behavioural Ecology and Evolutionary Genetics for helpful discussion. Dave Westneat is gratefully acknowledged for feedback and Sylvia Kuhn for lab work. The study was supported by the German Academic Exchange Service (DAAD), the Max Planck Society, and the International Max Planck Research School for Organismal Biology.

References

Araya-Ajoy, Y. G. & Dingemanse, N. J. 2014. Characterizing behavioural “characters”: an evolutionary framework. *Proceedings. Biological sciences / The Royal Society*, **281**, 20132645.

Bates, D., Mächler, M. & Bolker, B. 2014. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*,

Birkhead, T. 1991. Sperm depletion in the Bengalese finch, Lonchura striata. *Behavioral Ecology*, **4**, 267–275.

Birkhead, T., Atkin, L. & Møller, A. 1987. Copulation behaviour of birds. *Behaviour*, **101**, 101–138.

Birkhead, T. R., Briskie, J. V. & Moller, A. P. 1993. Male sperm reserves and copulation frequency in birds. *Behavioral Ecology and Sociobiology*, **32**, 85–93.

Forstmeier, W., Martin, K., Bolund, E., Schielzeth, H. & Kempenaers, B. 2011. Female

extrapair mating behavior can evolve via indirect selection on males. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 10608–13.

Fusani, L. 2008. Testosterone control of male courtship in birds. *Hormones and behavior*, **54**, 227–33.

Gelman, A. & Hill, J. 2007. *Data analysis using regression and multilevel/hierarchical models.* New York: Cambridge University Press.

Hadfield, J. D., Richardson, D. S. & Burke, T. 2006. Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework. *Molecular Ecology*, **15**, 3715–3730.

Halfwerk, W., Bot, S., Buikx, J., van der Velde, M., Komdeur, J., ten Cate, C. & Slabbekoorn, H. 2011. Low-frequency songs lose their potency in noisy urban conditions. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 14549–54.

Jennions, M. D. & Petrie, M. 2000. Why do females mate multiply? A review of the genetic benefits. *Biological reviews of the Cambridge Philosophical Society*, **75**, 21–64.

Johnsen, A., Carter, K. L., Delhey, K., Lifjeld, J. T., Robertson, R. J. & Kempenaers, B. 2011. Laying-order effects on sperm numbers and on paternity: comparing three passerine birds with different life histories. *Behavioral Ecology and Sociobiology*, **66**, 181–190.

Kempenaers, B., Verheyen, G., Van den Broeck, M., Burke, T., Van Broeckhoven, C. & Dhondt, A. 1992. Extra-pair paternity results from female preference for high-quality males in the blue tit. *Nature*,

Kempenaers, B., Borgström, P., Loës, P., Schlicht, E. & Valcu, M. 2010. Artificial night lighting affects dawn song, extra-pair siring success, and lay date in songbirds. *Current Biology : CB*, **20**, 1735–9.

Kokko, H. 2005. Mate guarding, male attractiveness, and paternity under social monogamy. *Behavioral Ecology*, **16**, 724–731.

Krist, M., Nádvorník, P., Uvírová, L. & Bureš, S. 2005. Paternity covaries with laying and hatching order in the collared flycatcher Ficedula albicollis. *Behavioral Ecology and Sociobiology*, **59**, 6–11.

Lebigre, C., Arcese, P. & Reid, J. M. 2013. Decomposing variation in male reproductive success: age-specific variances and covariances through extra-pair and within-pair reproduction. *The Journal of Animal Ecology*, **82**, 872–883.

Mace, R. 1987. The dawn chorus in the great tit paras major is directly related to female

fertility. *Naure*, **330**, 745–746.

Magrath, M. J. L., Vedder, O., van der Velde, M. & Komdeur, J. 2009. Maternal effects contribute to the superior performance of extra-pair offspring. *Current biology : CB*, **19**, 792–7.

Mayer, C. & Pasinelli, G. 2013. New support for an old hypothesis: density affects extra-pair paternity. *Ecology and Evolution*, **3**, 694–705.

Michl, G., Török, J., Griffith, S. C. & Sheldon, B. C. 2002. Experimental analysis of sperm competition mechanisms in a wild bird population. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 5466–70.

R Core Team. 2014. R: A language and environment for statistical computing.

Schlicht, E. & Kempenaers, B. 2013. Effects of social and extra-pair mating on sexual selection in blue tits (Cyanistes caeruleus). *Evolution; international journal of organic evolution*, **67**, 1420–1434.

Schlicht, L., Girg, A., Loës, P., Valcu, M. & Kempenaers, B. 2012. Male extrapair nestlings fledge first. *Animal Behaviour*, **83**, 1335–1343.

Schlicht, L., Valcu, M., Loes, P., Girg, a. & Kempenaers, B. 2014. No relationship between female emergence time from the roosting place and extrapair paternity. *Behavioral Ecology*, **25**, 650–659.

Webster, M., Pruett-Jones, S., Westneat, D. & Arnold, S. 1995. Measuring the effects of pairing success, extra-pair copulations and mate quality on the opportunity for sexual selection. *Evolution; international journal of organic evolution*, **49**, 1147–1157.

Westneat, D. F. & Stewart, I. R. K. 2003. Extra-pair Paternity in Birds: Causes, Correlates, and Conflict. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 365–396.

Westneat, D., Sherman, P. & Morton, M. 1990. The ecology and evolution of extra-pair copulations in birds. *Current Ornithology*,

Westneat, D. F., Stewart, I. R. K. & Hatch, M. I. 2009. Complex interactions among temporal variables affect the plasticity of clutch size in a multi-brooded bird. *Ecology*, **90**, 1162–1174.

Appendix 1

Table S1. Mixed effect model results of laying date effects on (a) a male's extra-pair fertilization opportunity and (b) extra-pair fertilizations success in relation to the fertility status of his social female. Results presented are assuming a female fertile period starting 5 days before the onset of laying to the day before the first egg was laid. We present the mode of the posterior distribution and in parenthesis lower and upper limits of the 95% credible interval.

Dependent variable	(a) Opportunity	(b) Extra-pair fertilization success
Fixed effects	β (95% CI)	β (95% CI)
Before fertile period	36.36 (31.03, 41.85)	-4.17 (-4.72, -3.57)
During fertile period	90.24 (85.03, 95.52)	-1.83 (-2.12, -1.52)
After fertile period	37.30 (32.03, 42.89)	-4.13 (-4.73, -3.54)
Day effect before fertile period	5.72 (5.33, 6.11)	0.19 (0.06, 0.33)
Day^2 effect before fertile period	0.05 (0.02, 0.09)	0.00 (-0.01, 0.00)
Day effect fertile period	0.24 (-0.14, 0.62)	-0.05 (-0.10, -0.01)
Day^2 effect fertile period	-0.32 -0.36, -0.28)	0.00 (-0.01, 0.00)
Day effect after fertile period	-6.30 (-6.69, -5.88)	-0.21 (-0.36, -0.07)
Day^2 effect after fertile period	0.27 (0.24, 0.31)	0.00 (-0.02, 0.01)
Random effects	σ^2 (95% CI)	σ^2 (95% CI)
$V_{Brood\ ID}$	0.00	0.00
$V_{Male\ ID}$	0.00	1.17 (1.02, 1.32)
$V_{Plot\text{-}Year}$	336.7 (328.2, 345.5)	0.23 (0.14, 0.32)
$V_{Residual}$	1383 (1321, 1451)	-

Table S2. Mixed effect model results for the effect of opportunity in a male's probability of extra-pair fertilization success in relation to his social female fertility stage. Presented are the results across a range of possible fertile period lengths. Fertile periods from 11 to 15 days are average lengths of the population, because these fertile periods include among-female variation in clutch size (see methods). Fertile period for 3 to 7 days do not include variation associated to clutch size differences, therefore all females in the population have the same value. We present the mode of the posterior distribution and in parenthesis lower and upper limits of the 95% credible interval.

	Fertile period length (days)					
Model	3	5	7	11	13	15
Fixed effects	β (95% CI)	β (95% CI)	β (95% CI)	β (95% CI)	β (95% CI)	β (95% CI)
Before fertile period	-2.79 (-3.17, -2.42)	-2.85 (-3.23, -2.47)	-2.97 (-3.32, -2.61)	-3.26 (-3.67, -2.84)	-3.36 (-3.81, -2.88)	-3.50 (-4.01, -3.01)
During fertile period	-2.06 (-2.50, -1.64)	-2.42 (-2.73, -2.11)	-2.52 (-2.84, -2.21)	-2.64 (-3.05, -2.25)	-2.60 (-3.00, -2.22)	-2.57 (-3.02, -2.13)
After fertile period	-3.24 (-3.65, -2.80)	-3.16 (-3.55, -2.73)	-3.15 (-3.60, -2.75)	-3.27 (-3.70, -2.85)	-3.13 (-3.55, -2.73)	-3.36 (-3.79, -2.90)
Opportunity before fertile period	0.64 (0.39, 0.89)	0.60 (0.38, 0.84)	0.64 (0.41, 0.87)	0.79 (0.52, 1.09)	0.82 (0.49, 1.12)	0.93 (0.58, 1.27)
Opportunity during fertile period	2.26 (1.41, 3.17)	1.54 (1.02, 2.08)	1.12 (0.67, 1.55)	0.93 (0.56, 1.29)	0.85 (0.53, 1.17)	0.85 (0.52, 1.18)
Opportunity after fertile period	0.83 (0.56, 1.11)	0.77 (0.52, 1.02)	0.78 (0.54, 1.03)	0.84 (0.55, 1.12)	0.84 (0.54, 1.13)	0.95 (0.56, 1.31)
Random effects	σ^2 (95% CI)	σ^2 (95% CI)	σ^2 (95% CI)	σ^2 (95% CI)	σ^2 (95% CI)	σ^2 (95% CI)
$V_{\text{Brood ID}}$	0.26 (0.24, 0.29)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	0.04 (0.03, 0.05)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)a
$V_{\text{Male ID}}$	0.41 (0.36, 0.47)	0.47 (0.41, 0.53)	0.40 (0.35, 0.46)	0.44 (0.39, 0.50)	0.34 (0.30, 0.39)	0.31 (0.27, 0.36)
$V_{\text{Plot-Year}}$	0.10 (0.06, 0.14)	0.14 (0.10, 0.19)	0.09 (0.06, 0.14)	0.10 (0.06, 0.14)	0.08 (0.05, 0.11)	0.10 (0.06, 0.14)

General Discussion

My thesis focuses on the adaptive causes and evolutionary consequences of phenotypic variation in traits that are repeatedly expressed throughout the lifetime of an individual (so-called 'labile' characters). The expression of labile characters is underpinned by processes acting at multiple levels and we were interested in the evolutionary implications of this multi-level nature. On the one hand, variation at the among-individual level arises because of differences between individuals in genes and environmental conditions. On the other hand, variation at the within-individual level is caused by an individual's attempt to adjust its phenotype across repeated expressions to match current environmental conditions. As a general aim, we aimed to study the causes of among- and within-individual variation in labile traits from an evolutionary perspective, and understand how these levels of variation were connected to the way populations may respond to selection. The thesis has two major components. The first is a conceptual and methodological part (chapters one to three) that aims to fully integrate the multi-level nature of labile phenotypes in the study of evolutionary characters, phenotypic plasticity, and social evolution. The second (chapters four and five) is the empirical component that aims to test evolutionary hypotheses about the sources of evolutionary constraint in the alternative fertilization strategies of male great tits as a worked examle. This latter part focuses on the social environment as a source of phenotypic variation and the role of labile characters in mediating the distinct strategies.

The first chapter focuses on how to define and statistically characterize “behavioral characters”. In this chapter we argue that behavioral characters (and other labile traits) can be usefully studied using the “evolutionary character concept” (Wagner 2001), while we emphasize the need to acknowledge that behavioral phenotypes are special because of their multi-level nature. One of the main messages of this chapter is that empiricists should test whether the behavioral measurements they collect are actually quantifying the behavioral characters they are attempting to study. If this is not done, researchers may arrive at the wrong conclusions about the role of a behavioral character in a particular biological process. We framed this chapter as a conceptual and methodological study, but exemplify its applicability by defining and quantifying the behavioral character, aggressiveness, in wild male great tits. In our study, we assayed several behaviors that great tits express during

simulated territorial intrusions. As a worked example of our proposed framework, we tested whether the patterns of variation and covariation at different levels suggested that all of the behaviors were expressions of the same behavioral character (aggressiveness). We found that not all of them did and that some were more reliable predictors than others. Therefore, this study informed us about the best measure of aggressiveness to incorporate in further empirical analyses in the second part of the thesis aimed at studying how variation in aggressiveness is related to the alternative fertilization strategies of male great tits. We hope that this study helps other researchers to better define and measure the behavioral characters they are interested in.

The second chapter is a methodological study where we proposed a sampling design with a corresponding structure of a certain statistical model, the “mixed-effect” model, that enables the study of multi-level variation in individual reaction norms. The main aim of this study was to provide empiricists with tools to test theory about the adaptive nature of repeatable vs. non-repeatable variation in phenotypic plasticity. While developing a method to estimate repeatability of plasticity, it became clear to us that phenotypic plasticity has a multi-level nature. We realized this because estimating repeatability of plasticity required the quantification of variation at the among- and within-individual level. Recently, the causes of among-individual variation in phenotypic plasticity have received increasing attention (e.g., (e.g., Wolf et al. 2008) while on the contrary the causes of within-individual variation have been largely neglected (Dingemanse & Wolf 2013). In this chapter, we show with a simulated example, that within-individual variation in phenotypic plasticity can arise if the plastic response to one environmental gradient varies in response to a second environmental gradient (Westneat et al. 2014a). We also show that novel patterns of multi-level variation can be revealed with our proposed methodology. For instance, we show how this approach enables the estimation of variation among individuals in reaction norm intercepts at different temporal scales. This was particularly useful for the empirical component of the thesis, because we later needed to partition variation in aggressive behavior of great tits at different temporal scales (chapter four). As part of chapter two, we also explored the performance of different sampling designs in terms of statistical power, precision, and accuracy to quantify multi-level variation in reaction norm components. We show that, in general, big sample sizes are required to accurately estimate variation in these parameters, but the sample size needed also greatly depends on the specifics of each study system. Therefore, we developed an R

simulation package to help researchers tailor these sample design issues to the particularities of their study system. Furthermore, we used this simulation package to determine the robustness of the statistical models applied to the data set used in our empirical studies (chapter four).

In the third chapter, we urge researchers to combine theory and methods developed in behavioral ecology and quantitative genetics (especially indirect genetic effects theory; e.g., Wolf et al. 1998) when studying labile (behavioral) characters in a social context. Phenotypes whose expression is affected by the social environment are sometimes called “interactive” phenotypes, because they are the result of the interaction between phenotypes of different individuals. These types of phenotypes are very common in nature and play an important role in population level processes. For instance, cooperative behaviors, agonistic behaviors, and even life-history traits (e.g., extra-pair reproduction) are only expressed in a social context. Quantitative geneticists have developed a strong theoretical and methodological framework to study these types of traits and how they may affect the evolutionary trajectories of populations (Wolf et al. 1998; McGlothlin et al. 2010). Despite the important role of labile characters in mediating the interactions between individuals and their social environments, the multi-level nature of the social environment has not been fully integrated into this framework. For example, empirical studies within this discipline investigate how aggressiveness of conspecific partners (“the social environment”) affects the aggressive behavior of a focal individual (Wilson et al. 2009). In this scenario, the social environment of one individual can vary due to phenotypic differences between social partners (e.g., among-individual variation in aggressiveness), but also because of phenotypic variation within a social partner due to phenotypic plasticity (e.g., within-individual variation in aggressiveness). The evolutionary consequences of the social environment will depend on the level of variation in the social environment that is affecting the phenotypic expression of the focal individual. Another important part in the evolutionary process where the multi-level nature of phenotypes comes into play is in the way individuals respond to the social environment. The interaction between an individual and its social environment is mediated by phenotypic plasticity (social responsiveness) and, as detailed in chapter two, plasticity can also vary among- and within-individuals. In chapter three we therefore detail how the multi-level nature of the social environment and of social responsiveness may be incorporated into the social evolution framework.

In chapter four, we study the sources of evolutionary constraint on alternative routes to fertilization success of male great tits. First, we considered male extra-pair and within-pair reproduction as interactive phenotypes that are affected by the phenotypes of both the male and the female member of the breeding pair. We showed that male fertilization strategies depend heavily upon the phenotype of his female mate. This particular social environment effect (as proposed in chapter three) should affect the evolutionary response to selection of male reproductive strategies (Wolf 2003; Brommer & Rattiste 2008) and could thus explain abundant among-individual variation and phenotypic stasis commonly observed in natural populations. This result highlights the importance of the social environment (here: female partners) as a source of variation that can have profound evolutionary consequences. Some of the most studied biological phenomena, like sexual selection, competition, and cooperation, are explicitly performed in social contexts and mediated by labile traits. Our study is an empirical example of the importance of the social environment as a source of phenotypic variation in the expression of labile traits.

In chapter four we also studied trade-offs between male alternative fertilization routes from the perspective of life-history theory (Noordwijk & Jong 1986; Stearns 1989). This approach required the decomposition of among- and within-individual covariance patterns between different fertilization routes to test specific hypotheses about the processes causing covariation at each levels We showed that among-male trade-offs between within-pair and extra-pair reproduction are an additional explanation for the existence of among-individual variation in traits so closely linked to fitness. In chapter five, we corroborated this idea by studying whether within-pair fertilizations and extra-pair fertilizations occur at the same time, allowing for the possibility of a trade-off between the two. We found that as a result of spill-over effects of male within-pair behavior on his extra-pair fertilization behavior, both within- and extra-pair fertilizations indeed occurred at the same time. Therefore, investing in behaviors that facilitate extra-pair fertilizations may constrain a male's ability to secure within-pair fertilizations and vice-versa. This result is consistent with our finding that there is indeed a trade-off between extra-pair and within-pair reproduction in this species. The results in chapters four and five highlight the importance of acknowledging that trade-offs between life history traits of an individual can result from processes acting at different levels but also in different individuals.

Biological hypotheses and statistical models

One of the underlying themes of this thesis is the evolutionary implications of the multi-level nature of labile characters. To study this phenomenon, we combined theory and methods developed in quantitative genetics and behavioral ecology to test hypotheses about the adaptive causes and evolutionary consequences of phenotypic variation at the among- and within-individual levels. On the one hand, quantitative geneticists have developed theory and methods to partition phenotypic variation at multiple levels, while on the other, behavioral ecologists have developed theory about the ecological conditions where adaptive among- and within-individual variation should evolve (Dingemanse & Dochtermann 2014). We connected these two fields of evolutionary biology by using the mixed-effect modeling framework (Wilson et al. 2010; Dingemanse & Dochtermann 2013) as a statistical tool for practical purposes (e.g., estimation of parameters), but also for conceptual purposes (Westneat et al. 2014b). We used the statistical "phenotypic equation" (Nussey et al. 2007) as the bridge between our biological hypotheses and the parameter estimates derived from mixed-effect models. This allows for an unambiguous understanding of the biological processes generating phenotypic variation and how to quantify them. In this part of the discussion, the statistical phenotypic equation (Nussey et al. 2007) will be used as a backbone to discuss the generalities and underlying assumptions of the studies that were part of the thesis.

Below is a simple version of the phenotypic equation used to detail the multi-level nature of labile characters.

$$y_{ij} = (\beta_0 + ind_{0j}) + e_{0ij} \qquad \text{Eq. 1}$$

$$[e_{0ij}]\, N(0, \Omega_e) \qquad \Omega_e = [V_{e_0}] \qquad \text{Eq. 1.2}$$

$$[ind_{0j}]\, N(0, \Omega_{ind}) \qquad \Omega_{ind} = [V_{ind_0}] \qquad \text{Eq. 1.3}$$

Here y_{ij} is the expression of a labile character at instance *i* of individual *j*. β_0 is the population mean for the labile character. Each individual has an intrinsic value, which is represented as a deviation from the population mean. An individual's intrinsic value can also be understood as its mean phenotypic value. This intrinsic value is here assumed to come

from a normal distribution with a mean of zero and a variance estimated from the data (i.e., the "among-individual" variance; V_{ind_j}). e_{0ij} represents the deviation of observation *i* from the intrinsic value of individual *j* and is here also assumed to come from a normal distribution with a mean of zero and a variance to be estimated from the data (i.e., within-individual variance; $V_{e_{ij}}$). Within-individual variation is caused by the individual's plastic response to (known or unknown) environmental variables (assuming no measurement error). On the other hand, among-individual variation is caused by genetic or environmental differences that have permanent effects on an individual's phenotype (permanent environmental effects).

Labile characters are different from fixed characters because the expression of labile characters is affected by the environment in two ways: via irreversible plasticity causing among-individual variation (permanent environmental effects), and also via reversible plasticity causing within-individual variation (temporary environmental effects). In equation 1, we explicitly refer to the temporary environmental effects (e_{0ij}) causing within-individual variation $V_{e_{ij}}$. By extending equation 1, we can model the environmental effects causing differences between individuals by partitioning an individual's intrinsic value into its genetic and environmental components.

$$ind_{0j} = a_{0j} + pe_{0j} \qquad \text{Eq. 2}$$

$$\left[a_{0j}\right] N\left(0, \Omega_a\right) \qquad \Omega_a = \left[V_{a_0}\right] \qquad \text{Eq. 2.2}$$

$$\left[pe_{0j}\right] N\left(0, \Omega_{pe}\right) \qquad \Omega_{pe} = \left[V_{pe_0}\right] \qquad \text{Eq. 2.3}$$

Here the intrinsic value of each individual: (ind_{0j}) partitioned into the contribution of additive genetic effects (breeding value; a_{0j}) and the influence of permanent environmental effects (pe_{0j}). For simplicity, we do not include genetic dominance. Additive genetic and permanent environmental effects are also assumed to come from a normal distribution with a mean of zero and a variance estimated from the data (additive genetic variance; V_{a_0} and permanent environmental variance; V_{pe_0}, respectively). One of the main messages of my thesis is the importance of embracing the multi-level nature of labile characters when trying to understand

the adaptive causes and evolutionary consequences of variation in labile traits. Therefore, we highlighted the importance of acknowledging that variation and covariation among labile traits can be due to among-individual and/or within-individual processes. In our empirical studies we specifically test hypotheses that required decomposition into within- versus among-individual variation. However, we did not further partition the among-individual variation into its genetic versus environmental components. In the next section, We discuss how this affects the evolutionary inferences that we can draw and some of the general assumptions that are made when working at the phenotypic level alone.

The phenotypic gambit

Evolutionary processes can be fully understood when both phenotypic and genetic data are available (Roff 1992). However, estimating genetic parameters is problematic in most natural systems and tests of evolutionary hypotheses are often done exclusively using phenotypic variation. This has the implicit assumption that phenotypes represents the underlying genetic architecture. This assumption has been coined the “phenotypic gambit” (Grafen 1984) and various studies have addressed its validity (Reusch & Blanckenhorn 1998; Waitt & Levin 1998; Dochtermann 2011). Cheverud (1988) was the first to provide support for this idea, when he compared genetic with phenotypic correlations and concluded that the two were sufficiently similar to justify making evolutionary inferences from phenotypic data alone. Since then, additional support has been found in a wide range of taxa and phenotypic traits (e.g., Reusch and Blanckenhorn 1998; Waitt and Levin 1998), but calls for caution about its generality have also been made, especially for traits with a large environmental component (Sheldon et al. 2003). Given that labile traits have such a high environmental component (V_{pe_0} and $V_{e_{ij}}$), it is necessary to be aware of how the assumptions of the phenotypic gambit apply to labile characters. In the next sections, we will explicitly discuss the phenotypic gambit in relation to labile characters and my thesis.

In chapter four, we were interested in the mechanisms maintaining variation in within-pair and extra-pair fertilization traits in male great tits. We intended to test mechanisms maintaining this variation because, according to Fisher's fundamental theorem (Fisher 1930), selection should deplete genetic variation in traits closely linked to fitness. Fisher was referring explicitly to genetic variation. With this in mind, we tried to separate variation caused by environmental effects. By quantifying among-individual variation (V_{ind_0}) in male fertilization

routes, we were able to separate the temporary environmental effects ($V_{e_{ij}}$), but not the variation associated to permanent environmental effects (V_{pe_0}). Recalling equations 1 and 2, we can explicitly show the assumptions we made in relation to the phenotypic gambit. Evolutionary studies on labile characters that assume phenotypic observations reflect the genetic underpinning make the implicit assumption that permanent and temporary environmental effects are similar to the genetic effects (the assumption is that $a_{0j} \approx pe_{0j} \approx e_{ij}$). In our study we actually accounted for the variation caused by temporary environmental effects, but we could not separate variation due to permanent environmental effects (we assumed that $a_{0j} \approx pe_{0j}$). This twist of the phenotypic gambit seems less problematic, because it relies on fewer assumptions. In a recent review Dochtermann and colleagues (2015) showed that, in behavioral traits, about 50 % of the variation between-individuals (V_{ind_0}) is underpinned by genetic variation (V_{a_0}). If we can generalize from the conclusions of this study, the among-individual variation in labile characters in our study should be underpinned by at least some genetic variation. Therefore, it seems valid to make qualitative evolutionary inferences when studying unpartitioned among-individual variation, but quantitative interpretations need to be done with care.

Multilevel covariation between traits

As mentioned in the previous section, it is important to distinguish the sources of variation in phenotypic traits in order to understand phenotypic evolution, but it is also very important to determine the sources of covariation between traits (Lande 1979; Sheldon et al. 2003; Dingemanse & Dochtermann 2013). In different chapters of my thesis, we partitioned the phenotypic correlation between traits into its among- and within-individual components to test specific evolutionary hypotheses about the processes causing covariation at each level. The contribution of among- vs. within-individual processes in the phenotypic correlation between traits is determined by the following equation (see Dingemanse & Dochtermann 2013):

$$r_{P_{0y}P_{0z}} = r_{ind_{0y}ind_{0z}} \sqrt{\left(\frac{V_{ind_{0y}}}{V_{ind_{0y}} + V_{e_{0y}}}\right)\left(\frac{V_{ind_{0z}}}{V_{ind_{0z}} + V_{e_{0z}}}\right)} + r_{e_{0y}e_{0z}} \sqrt{\left(\frac{V_{e_{0y}}}{V_{ind_{0y}} + V_{e_{0y}}}\right)\left(\frac{V_{e_{0z}}}{V_{ind_{0z}} + V_{e_{0z}}}\right)} \quad \text{Eq. 3}$$

Here $r_{P_{0y}P_{0z}}$, $r_{ind_{0y}ind_{0z}}$, and $r_{e_{0y}e_{0z}}$ represent the phenotypic, among-individual, and within-individual (residual) correlations respectively; $V_{ind_{0y}}$, $V_{ind_{0z}}$ are the between- individual variances and $V_{e_{0y}}$, $V_{e_{0z}}$ represent the within-individual variances for behaviors y and z respectively. The phenotypic correlation of two labile traits that have been measured repeatedly for the same individual is determined by: how the "intrinsic (mean) value" of each individual is correlated across traits (among-individual correlation; $r_{ind_{0y}ind_{0z}}$), and how deviations across instances from each individual's intrinsic values are correlated across traits (within-individual correlation; $r_{e_{0y}e_{0z}}$). This is then weighted by the geometric repeatability (roughly: the average repeatability across the two behaviors). The among-individual correlations are caused by the joint influences of genetic and permanent environmental effects acting on the two traits simultaneously (detailed below). The residual correlation is caused by pleiotropic effects of environmental variables with temporary effects on both traits, and, empirically, also by correlated measurement error. In some chapters of my thesis, we partitioned phenotypic correlations to test specific evolutionary hypotheses about the processes causing covariance at each of these hierarchical levels. For example, in chapter one we quantified the among- and within-individual correlations between behaviors expressed during aggressive interactions of great tits to determine if there was a single underlying latent variable that we could define as the behavioral character "aggressiveness". We expected that if the different agonistic behaviors were underpinned by the same latent variable, the patterns of correlation should be similar at the among- and within-individual levels. In chapter four, we also partitioned covariance between within-pair and extra-pair reproduction to determine the existence of resource allocation trade-offs between these alternative fertilization routes. Based on life-history theory, we expected a different pattern of correlations at the different levels, and therefore we partitioned covariation between these reproductive routes into the among- versus within-individual components. These two levels of correlation are, importantly, also affected by different processes which we could not explicitly disentangle in our studies. To address how such processes could have affected the inferences of our studies, we will first decompose the among-individual covariance between traits into its different components, and then we will do the same for the within-individual covariance.

Decomposing among-individual correlations

The processes contributing to the among-individual correlations can be shown by equation 4 (Dingemanse and Dochtermann 2014).

$$r_{ind_{0y}ind_{0z}} = r_{a_{0y}a_{0z}}\sqrt{\left(\frac{V_{a_{0y}}}{V_{a_{0y}}+V_{pe_{0y}}}\right)\left(\frac{V_{a_{0z}}}{V_{a_{0z}}+V_{pe_{0z}}}\right)} + r_{pe_{0y}pe_{0z}}\sqrt{\left(\frac{V_{pe_{0y}}}{V_{a_{0y}}+V_{pe_{0y}}}\right)\left(\frac{V_{pe_{0z}}}{V_{a_{0z}}+V_{pe_{0z}}}\right)} \quad \text{Eq. 4}$$

Here $r_{ind_{0y}ind_{0z}}$, $r_{a_{0y}a_{0z}}$, and $r_{pe_{0y}pe_{0z}}$ represent the among-individual, genetic, and permanent environment correlations respectively; $V_{a_{0y}}$ and $V_{a_{0z}}$ are the additive genetic variance and $V_{pe_{0y}}$, $V_{pe_{0z}}$ represent the permanent environmental variances for behaviors y and z respectively. The additive genetic correlation is caused by pleiotropic effects of genes and linkage disequilibrium (Lynch and Walsh, 1988). The permanent environmental correlation is caused by pleiotropic effects of environment variables that have permanent effects on both traits. In chapter one, our prediction was that correlations between behaviors that are functionally related should be the same at the among- and within-individual level. A further test of our idea would have been to determine whether the additive genetic correlation was similar to the permanent environmental and the within-individual correlations. If we had partitioned the among-individual correlation in this study, we could have further substantiated support for our predictions. In chapter 4, we found that there was an among-individual correlation between extra-pair paternity gain and within-pair paternity loss. We do not know whether this correlation is mediated by antagonistic pleiotropic effects of genes or permanent environmental effects. A genetic underpinning would imply that this correlation causes an evolutionary constraint between the fertilization routes due to the genetic architecture. If the among-individual correlation was caused by a permanent environmental correlation, it would imply that there is a resource allocation trade-off mediated by permanent environmental effects. This among-individual correlation most likely has contributions from both of these processes (genetic and permanent environment effects). Dochtermann (2011) has shown, for behavioral traits, that phenotypic correlations are at least of similar sign to additive genetic correlations. We can assume that if the phenotypic correlations reflect the genetic correlations, the among-individual correlation (as opposed to unpartitioned phenotypic correlations) should be a better predictor of the additive genetic correlations. Following

Dochtermann 2011), here we can (as above) qualitative inferences about the genetic underpinning of the among-individual correlations, but quantitative interpretations should be taken with care.

Decomposing within-individual correlations

Similar to among-individual correlations, within-individual correlations between traits can be underpinned by different processes. To show this we will partition the within-individual correlations ($r_{e_{0y}e_{0z}}$) into its different components (Brommer 2013).

$$r_{e_{0y}e_{0z}} = r_{te_{0y}te_{0z}} \sqrt{\left(\frac{V_{te_{0y}}}{V_{te_{0y}} + V_{mer_{0y}}}\right)\left(\frac{V_{te_z}}{V_{te_{0z}} + V_{mer_{0z}}}\right)} + r_{mer_y mer_{0z}} \sqrt{\left(\frac{V_{mer_{0y}}}{V_{te_{0y}} + V_{mer_{0y}}}\right)\left(\frac{V_{mer_{0z}}}{V_{te_{0z}} + V_{mer_{0z}}}\right)} \quad \text{Eq. 5}$$

Where $r_{e_y e_z}$, $r_{te_y te_z}$, and $r_{mer_y mer_z}$ represent the within-individual, temporary environment, and measurement error correlations respectively; V_{te_y} , V_{te_z} and, V_{mer_z} , V_{mer_z} represent the temporary environmental and measurement error variances for behaviors y and z respectively. From this equation, we can see that within-individual correlations can be caused by pleiotropic effects of unmeasured environmental variables and correlated measurement error. Within-individual correlations have, for a long time, been considered a nuisance mainly caused by correlated measurement error (Wilson et al. 2010; Brommer 2013). However, within-individual correlations are potentially the result of a very important biological process (e.g., Niemelä et al. 2015). To adaptively match the surrounding environment, individuals need to adjust several traits simultaneously to rapidly changing environments. Multiple traits need to respond not only to a changing environment, but also to various environmental gradients (Westneat et al. 2009). The within-individual correlation may capture this integrated multivariate plastic response to the multivariate environment (Westneat et al. 2009). This correlation level is of paramount importance because, to plastically adapt to the environment, functionally related traits should respond as a unit to environmental changes (Pigliucci 2001a). To be more specific, the level of integration in the reversible plasticity of different traits is captured by the temporary environmental correlation ($r_{te_y te_z}$). We made this argument explicitly in chapter one, where we posed that the within-individual correlation captures the integration of reversible plasticity in behaviors expressed during agonistic interactions. The

biological inferences that can be made from the within-individual correlations are determined by how it is affected by measurement error (Brommer 2013). Within-individual correlation can be biased upward due to correlated measurement error and downwards by uncorrelated measurement error. The degree of correlated measurement error will vary from situation to situation, but it is possible to account for its different sources. Experimental designs specifically tailored to separate measurement error are preferred in other to reduce its influence (Perktaş & Gosler 2010). It is also possible to statistically tease apart the effect of measurement error if the source is known (e.g., by modeling observer effects as we have done in chapter one). Within-individual correlations have received very little biological attention, but from an adaptive perspective they encode very important information about the integration of reversible plasticity of phenotypes. We think that studying variation among-populations and among-individuals in the level of integration of plasticity will increase our understanding of the evolutionary ecology of multivariate plasticity (Robinson & Beckerman 2013). This will require very high sample sizes in terms of number of individuals but also in terms of repeats within-individuals. New technologies my help to overcome the difficulty of collecting these type of data (Houle et al. 2010).

Multi-level variation in labile characters: adaptive causes

To conclude my thesis, we will discuss the different causes of phenotypic variation in labile characters from an adaptive and developmental perspective. The different components of phenotypic variation in labile characters arise in different stages of an individual's developmental time line (Figure 1, y axis). In each of these stages phenotypic variation is caused by different mechanisms. As mentioned in previous sections of the discussion, phenotypic variation in labile characters can arise due to genetic variation (Figure 1; V_{a_0} , blue bar), environmental variation mediated by irreversible plasticity (Figure 1; V_{pe_0} , green bar), and environmental variation mediated by reversible plasticity (Figure 1; V_{e_0} , red bar). The adaptive nature of variation generated by the different mechanisms will depend on the level of among- and within-individual variation in environmental conditions (Figure 4, x axis; detailed below). Among-individual variation in the environment is caused by differences between individuals in their environmental conditions (e.g., differences between territories). Within-individual variation in the environment are caused by temporal changes in the

environmental conditions that each individual experiences through his life (e.g., yearly variation in food availability).

Genetic variation arises by mutation and recombination, but the proportion of this variation resulting in adaptive among-individual differences may be caused by environment-dependent selection in the parental environment (Figure 1, blue bar). In other words, adaptive differences among individuals due to genetic variation may arise if natural selection favors different phenotypes depending on the environmental conditions, and the environmental differences are consistent across and within generations. For example, we can think about an heterogeneous environment that differs spatially in predator abundance. In this scenario, selection favors individuals that are more or less bold depending on the levels of predation risk. Boldness is a heritable trait and predator abundances are typically consistent in their spatial distribution over several generations. Therefore, in this particular example, adaptive genetic variation in boldness is expected to exist if there is among-individual variation in predation risk and predation risk is repeatable and predictable across and within generations.

Adaptive individual differences due to environmental effects, mediated by irreversible plasticity, are generated during an individuals' developmental phase (Figure 4, green bar). We define the developmental environment, broadly, as the environmental conditions during the time individual's are susceptible to environmental effects that will “fix” their phenotypes for life. The developmental phase stops once an individual's phenotype is canalized or crystallized. Therefore, given our definition of the developmental period, permanent environmental effects can only affect the phenotype in the developmental phase. Actually, what we refer as irreversible plasticity has also been coined developmental plasticity by other researchers (e.g., Stamps & Groothuis 2010). These types of environmental effects will be especially advantageous when early environmental cues can predict selective environments experienced by individuals later in life (Gabriel et al. 2005; Kuijper et al. 2014). Compared to adaptive genetic variation, permanent environmental effects offer individuals a way to adjust their phenotype to future environmental conditions when there is more information about the selective environment. Environmental cues about the “future environment” should be more reliable in the early environment of an individual than in the environment of its parents. Therefore, early environmental cues allow organisms to increase their match with their future (selective) environment (West-Eberhard 1989). Permanent environmental effects are particularly interesting when studying labile traits. From an adaptive perspective, why canalize

potentially plastic traits if adaptation to the environment can also be achieved by reversible plasticity? Adjusting to environmental conditions by reversible plasticity will allow individuals to change their phenotype if environmental conditions change. The existence of permanent environmental effects in labile traits, from an adaptive perspective, may imply that there are costs associated with relying partly or totally on reversible plasticity to cope with environmental conditions (DeWitt et al. 1998; Gabriel et al. 2005; Auld et al. 2010; Botero et al. 2014).

Reversible plasticity allows individuals to respond to current environmental conditions and therefore adapt to variation in their environment throughout their lives (Pigliucci 2001b). Environmental input mediated by reversible plasticity can cause both among- and within-individual variation in behavior. Unpredictable within-individual variation in the environment can only be assimilated by organisms via reversible plasticity (e.g., day to day variation in temperature), but individuals can also rely on reversible plasticity to cope with consistent among-individual variation in the environment (e.g., spatial variation in their life long territory). Therefore, within-individual variation in the environment will result in within-individual variation in behavior and among-individual variation in the environment will result in among-individual variation in behavior (Figure 4, red bar). Additionally, reversible plasticity can differ a great deal in speed and reversibility of change. At one extreme end of the spectrum, some traits may respond to immediate changes in the environment, while on the other end, some plastic changes might be relatively slow and may seem irreversible depending on the time scale of a study (Gabriel et al. 2005). Empirically, it is sometimes difficult to disentangle phenotypic variation arising from irreversible or reversible plasticity when environmental effects are long lasting or fixed during the life time of an individual. Despite the difficulty of disentangling these processes, it is important because these two types of processes can have very different evolutionary implications for populations (West-Eberhard 1989).

Understanding the different ways by which organisms can cope with environmental variation is at the intersection of most of today's biological disciplines (Pigliucci 2001b). Theoreticians interested in how individuals respond to environmental variation have studied the ecological conditions where reversible versus non-reversible plasticity should evolve (Gabriel et al. 2005; Botero et al. 2014).

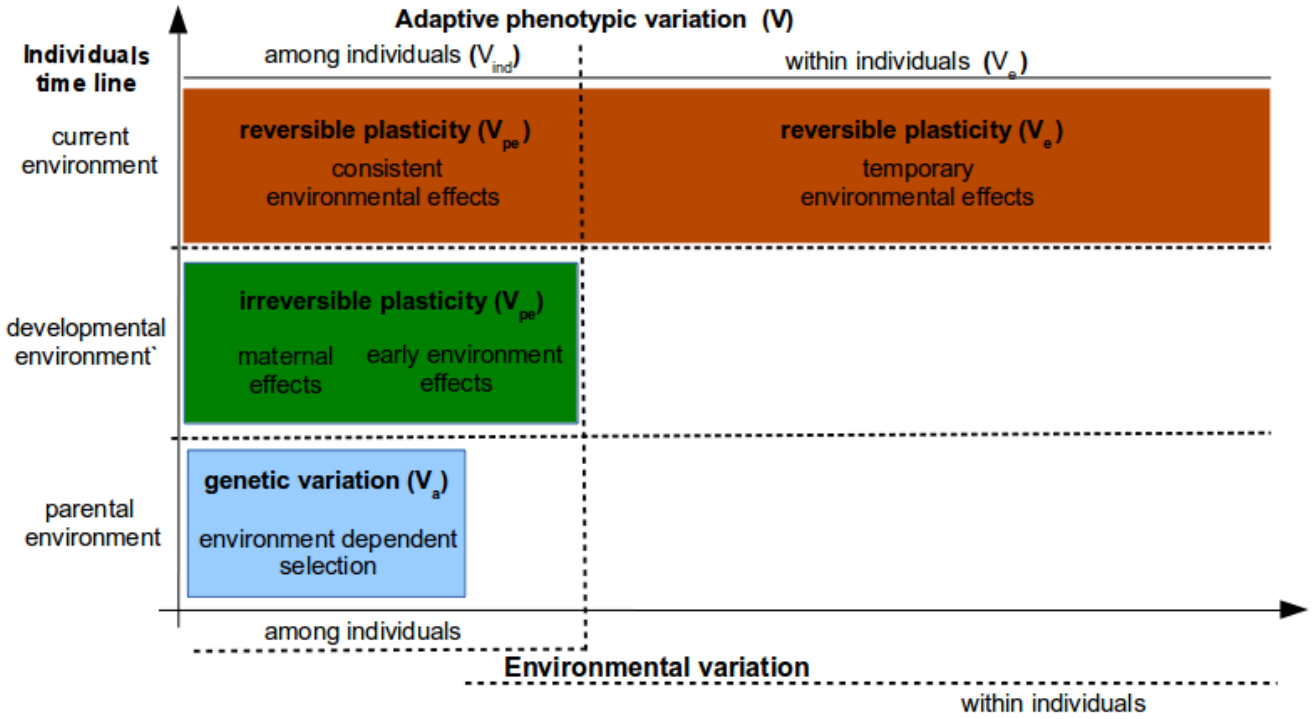

Figure 1. *Schematic representation of the different processes generating adaptive phenotypic variation in relation to the moment in an individual's time line when the environment affects phenotypic expression. The y-axis represents a discrete partition of an individual's time line (parental environment, developmental environment, and current environment). Genetic variation causing adaptive differences between individuals is generated in the environment experienced by parental generations. Variation caused by irreversible plasticity due to permanent environmental effects is generated during an individual's developmental stage. Variation caused by environmental effects due to reversible plasticity is generated by the current environment. On the x-axis, we depict the amount and type of environmental variation that will result in the different types of adaptive phenotypic variation. Genetic variation can only be an adaptive response to among-individual variation, while permanent environmental variation can also respond to among-individual variation in the environment and also to within-individual variation in the developmental environment. Irreversible plasticity can cope with environmental variation of both types. Environmental effects mediated by reversible phenotypic plasticity can result in both adaptive among- and within-individual phenotypic variation.*

However, this theoretical framework does not approach this problem from a quantitative perspective, as this methodological approach does not acknowledge that the phenotypic expression of labile characters is affected by different process but in different magnitudes. On the contrary, quantitative geneticists approach phenotypic variation in a quantitative way, by partitioning phenotypic variation into the different sources contributing to phenotypic expression (Lynch & Walsh 1998). Because quantitative geneticists focus on the genetic component of phenotypes, they are not as interested in studying the phenotypic variation caused by irreversible plasticity or reversible plasticity. Conversely, behavioral ecologists have been focusing on phenotypic variation caused by reversible plasticity for a long time (Krebs &

Davies 1997) and have more recently developed theoretical models about the adaptive causes of among-individual variation (Dingemanse & Wolf 2010). However, in their models behavioral ecologists often do not specify by which of the different processes (genetic variation, irreversible, or reversible plasticity) individuals are predicted to become different. We think one of the next challenges for evolutionary ecologists will be to integrate developments of these different fields to study phenotypic variation in a unified evolutionary framework.

Conclusion

Understanding the different processes generating adaptive phenotypic variation is important not only because it is a central theme in evolutionary ecology (Pianka 2011), but also because we need to understand how phenotypic variation at different levels helps individuals and populations to cope with the rapid environmental change induced by human activities (Thomas et al. 2004). The findings of this thesis call for the integration of different fields of biology to study the adaptive causes and evolutionary consequences of phenotypic variation in labile traits. This is a promising conclusion since one of the most successful ways to answer any scientific question is to integrate different fields of research (Kuhn 1996; Wilson 1999). On one hand, theoreticians working in behavioral ecology and other fields of evolutionary biology should develop models that generate quantitative predictions about the relative contributions of the different processes to phenotypic variation in a population. On the other hand, empiricists should use appropriate experimental designs and statistical tools to quantify the different sources of phenotypic variation when testing theoretical predictions about the adaptive nature of the different levels of variation. The aforementioned developments in these different fields of evolutionary ecology will allow them to be most effectively integrated. Charles Darwin conceived the idea of evolution by natural selection when he connected among- individual variation with among-species variation through the process of natural selection (Darwin 1859). He was not aware of genetic variation, but empirical and theoretical developments since the discovery of genetic inheritance have refined our understanding of the processes of evolution (Huxley 1943). Currently, the different ways by which individuals may cope with environmental variation during development and during adulthood are being incorporated in our understanding of the evolutionary process (West-Eberhard 2003). My thesis helped furthering this integration of the different processes

causing among- and within-individual variation into an evolutionary framework. Hopefully it will stimulate multidisciplinary research to increase our understanding of the evolutionary process.

References

Auld, J. R., Agrawal, A. & Relyea, R. a. 2010. Re-evaluating the costs and limits of adaptive phenotypic plasticity. *Proceedings. Biological sciences / The Royal Society*, **277**, 503–11.

Botero, C. A., Weissing, F. J., Wright, J. & Rubenstein, D. R. 2014. Evolutionary tipping points in the capacity to adapt to environmental change. *Proceedings of the National Academy of Sciences*, **112**, 201408589.

Brommer, J. E. 2013. On between-individual and residual (co)variances in the study of animal personality: are you willing to take the “individual gambit”? *Behavioral Ecology and Sociobiology*, **67**, 1027–1032.

Brommer, J. E. & Rattiste, K. 2008. “Hidden” reproductive conflict between mates in a wild bird population. *Evolution; international journal of organic evolution*, **62**, 2326–2333.

Cheverud, J. 1988. A comparison of genetic and phenotypic correlations. *Evolution*, **42**, 958–968.

Darwin, C. 1859. *The origin of species by means of natural selection.* London: Jhon Murray.

DeWitt, T., Sih, A. & Wilson, D. 1998. Costs and limits of phenotypic plasticity. *Trends in Ecology & Evolution*, **5347**, 77–81.

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

Dingemanse, N. J. & Dochtermann, N. A. 2014. Individual behaviour: behavioural ecology meets quantitative genetics. In: *Quantitative Genetics in the Wildld*, (Ed. by A. Charmantier, D. Garant, & L. E. B. Kruuk), pp. 54–67. Oxford University Press.

Dingemanse, N. J. & Wolf, M. 2010. Recent models for adaptive personality differences: a review. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **365**, 3947–58.

Dingemanse, N. J. & Wolf, M. 2013. Between-individual differences in behavioural plasticity within populations: causes and consequences. *Animal Behaviour*, **in press**, 1–9.

Dochtermann, N. A. 2011. Testing cheverud’s conjecture for behavioral correlations and behavioral syndromes. *Evolution; international journal of organic evolution*, **65**, 1814–20.

Dochtermann, N. A, Schwab, T. & Sih, A. 2015. The contribution of additive genetic variation to personality variation: heritability of personality. *Proceedings. Biological sciences / The Royal Society*, **282**, 20142201.

Fisher, R. A. 1930. *The genetical theory of natural selection.* The Clarendon Press.

Gabriel, W., Luttbeg, B., Sih, A. & Tollrian, R. 2005. Environmental tolerance, heterogeneity, and the evolution of reversible plastic responses. *The American Naturalist*, **166**, 339–53.

Grafen, A. 1984. Natural selection, kin selection and group selection. In: *Behavioral Ecology: An Evolutionary Approach*, 2nd edn.- (Ed. by J. R. Krebs & N. B. Davies), pp. 62–84. Oxford: Blackwell.

Houle, D., Govindaraju, D. R. & Omholt, S. 2010. Phenomics: the next challenge. *Nature reviews. Genetics*, **11**, 855–66.

Huxley, J. 1943. *Evolution: The Modern Synthesis.* Harper & brothers.

Krebs, J. R. & Davies, N. B. 1997. *Behavioural Ecology: An Evolutionary Approach.* Blackwell Science.

Kuhn, T. S. 1996. *The Structure of Scientific Revolutions.* University of Chicago Press.

Kuijper, B., Johnstone, R. A. & Townley, S. 2014. The Evolution of Multivariate Maternal Effects. *PLoS Computational Biology*, **10**, e1003550.

Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution*, **33**, 402–416.

Lynch, M. & Walsh, B. 1998. *Genetics and analysis of quantitative traits.* Sunderland, Mass.: Sinauer.

McGlothlin, J. W., Moore, A. J., Wolf, J. B. & Brodie, E. D. 2010. Interacting phenotypes and the evolutionary process. III. Social evolution. *Evolution; international journal of organic evolution*, **64**, 2558–2574.

Niemelä, P., Lattenkamp, E. Z. & Dingemanse, N. J. 2015. Personality-related survival and sampling bias in wild cricket nymphs. *Behavioral Ecology*,

Noordwijk, A. Van & Jong, G. De. 1986. Acquisition and allocation of resources: their influence on variation in life history tactics. *American Naturalist*, **128**, 137–142.

Nussey, D. H., Wilson, a J. & Brommer, J. E. 2007. The evolutionary ecology of individual phenotypic plasticity in wild populations. *Journal of Evolutionary Biology*, **20**, 831–44.

Perktaş, U. & Gosler, A. G. 2010. Measurement Error Revisited: Its Importance for the Analysis of Size and Shape of Birds. *Acta Ornithologica*, **45**, 161–172.

Pianka, E. R. 2011. *Evolutionary Ecology.* Eric R. Pianka.

Pigliucci, M. 2001a. Characters and environments. In: *The Character Concept in Evolutionary Biology*, (Ed. by G. Wagner), pp. 365–390. San Diego, CA: Academic Press.

Pigliucci, M. 2001b. *Phenotypic Plasticity: Beyond Nature and Nurture.* Johns Hopkins University Press.

Reusch, T. & Blanckenhorn, W. 1998. Quantitative genetics of the dung fly Sepsis cynipsea: Cheverud's conjecture revisited. *Heredity*, **81**, 111–119.

Robinson, M. R. & Beckerman, A. P. 2013. Quantifying multivariate plasticity: genetic variation in resource acquisition drives plasticity in resource allocation to components of life history. *Ecology letters*, **16**, 281–90.

Roff, D. A. 1992. *Evolution Of Life Histories: Theory and Analysis.* Springer.

Sheldon, B. C., Kruuk, L. E. B. & Merila, J. 2003. When environmental variation short-circuits natural selection. **18**, 207–209.

Stamps, J. & Groothuis, T. G. G. 2010. The development of animal personality: relevance, concepts and perspectives. *Biological reviews of the Cambridge Philosophical Society*, **85**, 301–25.

Stearns, S. 1989. Trade-offs in life-history evolution. *Functional ecology*, **3**, 259–268.

Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., Erasmus, B. F. N., De Siqueira, M. F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., Van Jaarsveld, A. S., Midgley, G. F., Miles, L., Ortega-Huerta, M., Peterson, T., Phillips, O. L. & Williams, S. E. 2004. Extinction risk from climate change. *Nature*, **427**, 145–8.

Wagner, G. P. 2001. *The Character Concept in Evolutionary Biology.* San Diego, CA: Academic Press-Elsevier Science.

Waitt, D. & Levin, D. 1998. Genetic and phenotypic correlations in plants: a botanical test of Cheverud's conjecture. *Heredity*, **80**, 310–319.

West-Eberhard, M. J. 1989. Phenotypic Plasticity and the Origins of Diversity. *Annual Review of Ecology and Systematics*, **20**, 249–278.

West-Eberhard, M. J. 2003. *Developmental Plasticity and Evolution.* OUP USA.

Westneat, D. F., Stewart, I. R. K. & Hatch, M. I. 2009. Complex interactions among temporal variables affect the plasticity of clutch size in a multi-brooded bird. *Ecology*, **90**, 1162–1174.

Westneat, D. F., Bókony, V., Burke, T., Chastel, O., Jensen, H., Kvalnes, T., Lendvai, Á. Z., Liker, A., Mock, D., Schroeder, J., Schwagmeyer, P. L., Sorci, G. & Stewart, I. R. K. 2014a. Multiple aspects of plasticity in clutch size vary among populations of a globally distributed songbird. *Journal of Animal Ecology*, **83**, 876–887.

Westneat, D. F., Wright, J. & Dingemanse, N. J. 2014b. The biology hidden inside residual within-individual phenotypic variation. *Biological Reviews*, n/a–n/a.

Wilson, E. O. 1999. *Consilience: The Unity of Knowledge.* Vintage Books.

Wilson, A. J., Gelin, U., Perron, M.-C. & Réale, D. 2009. Indirect genetic effects and the evolution of aggression in a vertebrate system. *Proceedings. Biological sciences / The Royal Society*, **276**, 533–41.

Wilson, A. J., Réale, D., Clements, M. N., Morrissey, M. M., Postma, E., Walling, C. a, Kruuk, L. E. B. & Nussey, D. H. 2010. An ecologist's guide to the animal model. *The Journal of Animal Ecology*, **79**, 13–26.

Wolf, J. B. 2003. Genetic architecture and evolutionary constraint when the environment contains genes. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 4655–4660.

Wolf, J. B., Brodie, E. D., Cheverud, J. M., Moore, A. J. & Wade, M. J. 1998. Evolutionary consequences of indirect genetic effects. *Trends in Ecology & Evolution*, **13**, 64–69.

Wolf, M., van Doorn, G. S. & Weissing, F. J. 2008. Evolutionary emergence of responsive and unresponsive personalities. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 15825–30.

Acknowledgements

I would like to express my profound gratitude to my supervisor, Niels Dingemanse, whose patience, knowledge, and understanding shaped the course of my PhD. As a person and scientist, he has been a source of support and motivation during these years.

I am also very grateful to Bart Kempenaers for the many insightful discussions and support during my PhD.

I also want to thank all past and present members of the Evolutionary Ecology of Variation group for the stimulating feedback and their help in the collection of the data. I would like to especially thank Robin Abbey-Lee, Kim Mathot, Alexia Mouchet, Ariane Mutzel, Marion Nicolaus, Maria Moirón, Anne-Lise Olsen, Jan Wijmenga, and Erica Stuber for the amazing time I had doing fieldwork. I would also like to thank Kim Mathot for the time she took to discuss with me and her willingness to help whenever she could. Sylvia Kuhn for her meticulous work in the genotyping and parentage analysis of the great tits. Mihai Valcu, Pim Edelaar, Jon Wright, Ned Dochtermann, Dave Westneat, Holger Schielzeth, and Wolfgang Forstmeier for the time they spent discussing and sharing their knowledge with me. Thanks to Francesca Santostefano for all the matrices we discussed, to Petri Niemelä, Martin Bulla, Maria Moirón, and Ulrich Knief for many interesting discussions. The members of the department of Behavioural Ecology and Evolutionary Genetics for all their help during these years and the International Max Planck Research School for Organismal Biology, the Max Planck Institute for Ornithology, and the DAAD for their financial and logistical support.

For their help in the various aspects of life where one needs help from his friends, I want to thank Ulrich Knief, Rene Quispe, and Camilla Villavicencio. I also want to thank Hannah Fried-Petersen for all her help and patience, day and night. Last but not least, I want to thank my family for their unconditional support.

Author Contributions

Chapter 1 Characterizing behavioural “characters”: An evolutionary framework.
YA-A and NJD conceived the study and collected the data. YA-A wrote the manuscript and NJD constributed substantially to revisions.

Chapter 2 Multi-level analysis of reaction norms: an approach to estimate short-term, long-term and reaction norm repeatability.
YA-A and NJD conceived the study. YA-A and NJD analyzed the data. YA-A and NJD wrote the manuscript with input from KJM.

Chapter 3 Interacting personalities: behavioral ecology meets quantitative genetics.
YA-A and NJD conceived the study. NJD wrote the manuscript with major input from YA-A.

Chapter 4 Evolutionary constraints in the alternative routes to fertilization success of male great tits: trade-offs and indirect female effects.
YA-A, NJD and BK conceived the study. YA-A, KJM, AM, MN, JW and NJD collected the data. YA-A, SK, and BK performed the paternity analysis. YA-A and NJD analyzed the data with the input of BK. YA-A and NJD wrote the manuscript; all other authors contributed substantially to revisions.

Chapter 5 Timing of extra-pair fertilizations: within-pair fertilization trade-offs or pair synchrony spill-overs
The authors jointly conceived the study. YA-A and NJD collected the data. YA-A and BK analyzed the data with the input of NJD. YA-A, BK wrote the manuscript with substantial input from NJD.

Niels J. Dingemanse	Yimen G. Araya Ajoy
Doktorvater	Doktorant

Addresses of Co-Authors

Niels J. Dingemanse
Behavioural Ecology, Department Biology II
Ludwig-Maximilians University of Munich,
Planegg-Martinsried, Germany

Evolutionary Ecology of Variation Research Group
Max Planck Institute for Ornithology,
Seewiesen, Germany
Email: ndingemanse@orn.mpg.de

Bart Kempenaers
Department of Behavioral Ecology and Evolutionary Genetics
Max Planck Institute for Ornithology
Seewiesen, Germany
Email: b.kempenaers@orn.mpg.de

Sylvia Kuhn
Department of Behavioral Ecology and Evolutionary Genetics
Max Planck Institute for Ornithology
Seewiesen, Germany
Email: skuhn@orn.mpg.d

Kimberley J. Mathot
Marine Ecology Department
Royal Netherlands Institute for Sea Research (NIOZ)
Den Burg, Texel, the Netherlands
Email: kjmathot@gmail.com

Alexia Mouchet

Evolutionary Ecology of Variation Research Group

Max Planck Institute for Ornithology,

Seewiesen, Germany

Email: amouchet@orn.mpg.de

Ariane Mutzel

Department of Biology

University of Kentucky

Lexington, Kentucky, United States

Email: ariane.mutzel@gmx.de

Marion Nicolaus

Department of Molecular Biology and Biochemical Engineering,

University Pablo de Olavide

Sevilla, Spain

Email: nicolaus.marion@gmail.com

Jan Wijmenga

Evolutionary Ecology of Variation Research Group

Max Planck Institute for Ornithology,

Seewiesen, Germany

Email: janwijmenga@gmail.com

Curriculum Vitae

Yimen Gerardo Araya Ajoy
Evolutionary Ecology of Variation Group
Max Planck Institute for Ornithology
E-mail: yimencr@gmail.com

Nationality: Costa Rican.

Current activity

PhD student, International Max Planck Research School for Organismal Biology.

Education

Master in Biological Sciences (Honors), University of Costa Rica (UCR), 2009.
Bachelor in Biology, University of Costa Rica (UCR), 2005.

Awards and scholarships

Deutscher Akademisher Austauschdienst Dienst, Doctoral Fellowship, 2012-2014.
International Max Planck Research School for Organismal Biology, Doctoral Fellowship, 2010.
Grant to attend conference, Costa Rican Council of Sciences and Technology, 2008.
Research exchange scholarship, DAAD-Ulm University, Germany, 2007.
Research fellowship, Organization of Tropical Studies (OTS), 2006.
Research Experience for Undergraduates, OTS 2005.
Raven Award program, Cornell University, 2005.

Academic and research positions

Teacher, Workshop on Statistical Approaches to Multivariate Phenotypes, International Max Planck Research School for Organismal Biology (Germany) and Center for Biodiversity Dynamics (Norway), 2015.
Teacher, Behavioral Ecology meets Quantitative Genetics, Ludwig-Maximilians-University of Munich (LMU), 2013-2014.
Teaching assistant, Experimental Behavioural Ecology, LMU, 2013.

Teaching assistant, Tropical Ecology and Conservation course from the Council of International Exchange Education (CIEE) 2009-2010.

Teaching assistant, Ecology, University of Costa Rica (UCR), 2008.

Teaching assistant, Introduction to Biology, University of Costa Rica (UCR), 2008.

Research assistant, Department of Experimental Ecology, Ulm University, Germany, 2007.

Teaching assistant, Bioacoustics Course, UCR, 2006.

Research assistant, Zoology Museum, UCR, 2005-2006.

Publications

Dingemanse, N. J. & **Araya-Ajoy, Y. G.** 2014**.** Interacting personalities: behavioural ecology meets quantitative genetics. *Trends in Ecology & Evolution*, 30, 88–97.

Nicolaus, M., Mathot, K. J., **Araya-Ajoy, Y. G.,** Mutzel, A., Wijmenga, J. J., Kempenaers, B. & Dingemanse, N. J. 2014. Does coping style predict optimization? An experimental test in a wild passerine bird. *Proceedings. Biological Sciences / The Royal Society*, 282, 20142405.

Araya-Ajoy, Y. G. & Dingemanse, N. J. 2014. Characterizing behavioural “characters”: an evolutionary framework. *Proceedings. Biological sciences / The Royal Society*, 281, 20132645.

Mathot, K. J., Nicolaus, M., **Araya-Ajoy, Y. G.,** Dingemanse, N. J. & Kempenaers, B. 2014**.** Does metabolic rate predict risk-taking behaviour? A field experiment in a wild passerine bird. *Functional Ecology*, 239–249.

Stuber, E. F., **Araya-Ajoy, Y. G.,** Mathot, K. J., Mutzel, a., Nicolaus, M., Wijmenga, J. J., Mueller, J. C. & Dingemanse, N. J. 2013. Slow explorers take less risk: a problem of sampling bias in ecological studies. *Behavioral Ecology*, 24, 1092–1098.

Araya-Ajoy, Y. G., Chaves-Campos, J., Kalko, E. K. V & Dewoody, J. A. 2009. High-pitched notes during vocal contests signal genetic diversity in ocellated antbirds. *PloS one*, 4, e8137.

Chaves-Campos, J., **Araya-Ajoy, Y. G.**, Lizana-Moreno, C. a & Rabenold, K. N**.** 2009. The effect of local dominance and reciprocal tolerance on feeding aggregations of ocellated antbirds. *Proceedings. Biological sciences / The Royal Society*, 276, 3995–4001.

Manuscripts submitted and in preparation

Araya-Ajoy, Y.G., Mathot, K. J., Dingemanse, N.J. Repeatable reaction norms and the analysis of short- versus long-term repeatability. Methods in Ecology and Evolution (2nd submission).

Araya-Ajoy, Y. G., Kempenaers, B., Kuhn, S., Mathot, K. J., Mouchet, A., Mutzel, A., Nicolaus A., Wijmenga, J., Dingemanse N. J. Evolutionary constraints in the alternative routes to fertilization success of male Great tits: trade-offs and indirect female effects. (in preparation).

Araya-Ajoy, Y. G., Dingemanse N. J., Kempenaers, B. Timing of extra-pair fertilizations: within-pair fertilization trade-offs or pair synchrony spill-overs. (in preparation).

Journal reviewer

Animal Behaviour, Behavioral Ecology, Behavioral Ecology and Sociobiology, Ethology, Journal of Comparative Psychology, Journal of Evolutionary Biology.

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt ist.

München, den ....

(Unterschrift)

Erklärung

Hiermit erkläre ich, *

☐ dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist.

☐ dass ich mich anderweitig einer Doktorprüfung ohne Erfolg **nicht** unterzogen habe.

☐ dass ich mich mit Erfolg der Doktorprüfung im Hauptfach .....

und in den Nebenfächern .....

bei der Fakultät für ..... der .....

(Hochschule/Universität)

unterzogen habe.

☐ dass ich ohne Erfolg versucht habe, eine Dissertation einzureichen oder mich der Doktorprüfung zu unterziehen.

München, den..... .....

(Unterschrift)

*) Nichtzutreffendes streichen