An interdisciplinary perspective on personality

Dissertation

for the award of the degree

Doctor rerum naturalium (Dr. rer. nat.)

of the Georg-August University Göttingen

in the PhD programme Behavior and Cognition

of the Georg-August University School of Science (GAUSS)

submitted by

Christoph Julian von Borell du Vernay

born in Bad Oldesloe, Germany

Göttingen, 2022

Thesis Committee

Lars Penke, Georg-August-University Göttingen

Oliver Schülke, Georg-August-University Göttingen

Niels Dingemanse, Ludwig-Maximilians-University Munich

Reviewing Members of the Examination Board

First Reviewer: Lars Penke

Second Reviewer: Niels Dingemanse

Further Members of the Examination Board

Tanya Behne, Georg-August-University Göttingen

Margarethe Boos, Georg-August-University Göttingen

York Hagmeyer, Georg-August-University Göttingen

Oliver Schülke, Georg-August-University Göttingen

Date of Oral Examination

29.09.2022

Table of Contents

Preface

This dissertation is not a cumulative, publication-based dissertation, but follows it in form. It includes three manuscripts, two of which have been published and the third being in preparation for submission. All non-publication based parts of the dissertation were written by myself. Assistance of co-authors was provided in the included publications, but they were all written by myself in accordance with the scientific and examination regulations. All sources have been quoted.

I provide the manuscripts in the following order in the appendix.

von Borell, C., Dingemanse, N., Westneat, D., Denissen, J. & Penke, L. (2022). Towards a future of interdisciplinary personality research – exchanging ideas between personality psychology and behavioural ecology. In preparation.

von Borell, C., Weiss, A. & Penke, L. (2019). Developing individual differences in primate behavior: The role of genes, environment and their interplay. Behavioral Ecology and Sociobiology, 73, 20.https://dx.doi.org/10.1007/s00265-019-2633-8.

von Borell, C., Kordsmeyer, T., Gerlach, T. & Penke, L. (2019). An integrative study of facultative personality calibration. Evolution and Human Behavior, 40, 235 – 248. https://dx.doi.org/10.1016/j.evolhumbehav.2019.01.002.

Short summary

In this dissertation I approach the question of why individuals differ from in each other in their behavioural dispositions from an interdisciplinary perspective that builds upon previous work from the disciplines of psychology and behavioural ecology. I review terminological and methodological differences between these two disciplines, integrate findings on ontogenetic development of personality differences in primates with a focus on aspects of behaviour genetics, and provide empirical results that allow a differentiated evaluation of the hypothesis that personality differences between humans are the result of a calibration to differences in physical appearance.

Kurze Zusammenfassung

In meiner Doktorarbeit nähere ich mich aus einer interdisziplinären Perspektive der Frage warum sich Individuen in ihren Verhaltensdispositionen unterscheiden und beziehe mich dabei auf Arbeiten aus den Disziplinen der Psychologie und Verhaltensökologie. Ich diskutiere terminologische und methodische Unterschiede zwischen diesen Disziplinen, integriere Ergebnisse zur ontologischen Entwicklung von Persönlichkeitsunterschieden bei Primaten mit Fokus auf verhaltensgenetischen Aspekten und stelle empirische Ergebnisse vor, welche eine differenzierte Evaluation der Hypothese erlauben, das Persönlichkeitsunterschiede zwischen Menschen das Resultat einer Kalibrierung auf Unterschiede in ihrer physischen Erscheinung seien.

1. Introduction

Across species we can observe that individuals differ from each other in their behavioural propensities. Fish may differ in how reluctant they are to explore the open water, monkeys may differ in how calm they react in the face of a predator, humans may differ in how talkative they are in a new-formed group, to name a few examples. If such differences are persistent over time, they are commonly labelled personality or personality traits. In psychology, personality traits are usually conceptualised as global and temporally stable dispositions ("dimensions"), which affect multiple related behaviours, feelings, and thoughts (Jackson et al., 2010). Personality denotes the uniqueness of an individual based on it's mixture of values in all of these traits (Allport & Allport, 1921; Uher, 2008, 2011). Biologists, especially behavioural ecologists, also use the term "personality" to describe individual differences in the behaviour of non-human animal species (to whom I will subsequently refer to as "animals" for reason of brevity). Personality is in behavioural ecology however not an individual-centred concept and usually not perceived to be structured in global dimensions. Rather, personality is defined as between-individual variance in a given behaviour in the average environment (Dingemanse & Wright, 2020).

Although the aspect of temporal stability is sometimes stressed in the definition of personality, it is not a phenomenon without dynamics. Two aspects are not to be confused here. First, even if individuals are highly stable in their behaviour on average ("have personality"), they still will adjust their behaviour to specific situations or demands of their environment. This kind of adjustment is termed plasticity. A child that usually plays calm and relaxed on it's own may still demand for it's parents when something unexpected happens. Second, the average tendency of how an individuals behaves ("personality") may change during ontogeny, this is what we call ontogenetic personality development. Causes for such personality development are, for example, individual maturation or more permanent environmental change that leads to novel challenges that individuals face in their lifetime (Thompson, 1998; Orr, 2009; Stamps & Groothuis, 2010). This can be imagined by picturing the development of an easily distressed infant who turns into a relaxed grown-up or vice versa. In addition, we can also consider evolutionary dynamics that affect behavioural variation over many generations of individuals. From the logic of evolutionary selection, maladaptive dispositions are, depending on their severity, unlikely to be passed on to following generations,

because individuals that carry them have lower chances of survival or reproduction. Hence, behavioural dispositions of individuals either need to be adaptive or without much effect on the fitness of an individual to be passed on to another generation (Penke, Denissen, & Miller, 2007). The term *fitness*, in an evolutionary sense, summarizes the survival and reproductive success of organisms (or populations / species), meaning to what proportion their genotype (or phenotype) is represented in the next generations (Orr, 2009). The fitness of individuals depends on how well they are *adapted* to the environment they live in. Formally, such evolutionary perspective can be distinguished between past and present, that is, how did behavioural variation develop historically ("phylogeny") and what is the current utility (also termed "survival value" or "adaptive significance") of behavioural variation (Bateson & Laland, 2013; Nesse, 2013; Tinbergen, 1963).

Since the study of personality is an interdisciplinary subject of research, a naïve view could be that research findings from different disciplines can be simply combined to form a broad understanding of individual differences in behaviour. In fact, there have been attempts to link these literatures and stimulate research that can be transferred from one discipline to the other. However, despite these attempts, research within behavioural ecology and psychology have developed rather separately and still are not very well connected (Nettle & Penke, 2010).

Personality psychologists have invested many efforts to understand how individual differences between humans develop throughout the lifetime (ontogenetic perspective). By now, we have a comparatively solid understanding of how and when human personality tends to stabilize or change over different developmental stages (i.e., infancy, childhood, adolescence, adulthood) and how this is tied to genetic differences between individuals. We know that differences which can be compared across the lifespan are moderately stable, which, as the other side of the coin, also means that they underlie moderate change (Caspi, Roberts, & Shiner, 2005). Behavioural ecologists have not uncovered ontogenetic pathways of personality development as conclusively yet (Stamps & Groothuis, 2010). There are some findings about short-term vs. long-term stability of behavioural differences (e.g., Boulton et al., 2014) and also some studies that compare personality differences across different stages of animal development, especially in primates (e.g., von Borell, Kulik, & Widdig, 2016). However, for most species other than humans, there is

not much known how genetic differences and the environment interplay and shape personality during the lifetime of individuals that can be summarized to a general ontogenetic pattern.

If we consider the understanding of evolutionary processes behind the emergence of personality, differential research investments between psychology and biology are reversed. While some evolutionary psychologists have tried to understand human personality from an evolutionary perspective (that is, how personality is selected for across generations), the literature is dominated by the study of ontogenetic and physiological causes of personality differences (Penke, Denissen, & Miller, 2007; Lukaszewski 2013), which we can subsume to proximate explanations (Tinbergen, 1963). In behavioural ecology, however, understanding the evolutionary causes of personality, that is finding ultimate explanations, is a major research objective (Simmons, 2014). Research of behavioural ecologists showed, for example, that behavioural variation among individuals is adaptive under conditions of temporarily fluctuating environments (Dingemanse, Both, Drent, & Tinbergen, 2004) or across different environments that differ in predation risk (Giles & Huntingford, 1984).

Under the premise that proximate and evolutionary forces in personality development are comparable across humans and animals, one might consider to combine results on ontogenetic development from humans with results on evolutionary aspects and function from animals to understand personality dynamics across different levels. As I will show below, theoretical concepts of evolution and behavioural development can be transferred quite well across disciplines. For empirical results such transfer does, however, not go as smoothly. This is because studies of psychologists and behavioural ecologists often differ in their methodology of data collection, data aggregation and data analysis.

Before I summarize the content of the three manuscripts included in this thesis, I will provide an overview of background literature for the main topics of these manuscripts. I will first turn to ontogenetic development of personality, then go on to present fundamentals of evolutionary genetics of personality variation and then introduce methodological aspects entailed in the study of personality across disciplines.

1.1 Ontogenetic development of personality

Using the word "personality" as descriptor of individual differences in behaviour is often tied to the premise that these differences sustain a certain temporal stability. If we follow up upon this, how much "development" or change in a behavioural disposition is there to expect throughout the lifetime of individuals? Parents or caregivers of humans and animals may ask this question to get a sense of what is coming at them in the future. Ecologists may wonder to what extent animals can adapt their behaviour to changes in the environment or whether such change affects younger and older animals differently. Critics of the concept of personality may ask for a threshold upon which we consider behavioural variation a personality trait (e.g., Beekman & Jordan, 2017). Studies of personality development aim at answering such questions on lifetime stability and change in behaviour. Before I touch empirical findings on this matter, I will briefly explain how researchers of different disciplines quantify "stability" of behaviour or more global personality dimensions.

Verification of behavioural stability is usually tied to estimating the correlation between two or more behavioural measurements distributed across time, using, for example, the product-moment correlation coefficient r (for a correlation across two measurement occasions) or the intra-class correlation coefficient ICC (for more than two measurement occasions). In behavioural ecology the latter is termed coefficient of repeatability R (Nakagawa & Schielzeth, 2010; see manuscript I for more details). There is no general rule or threshold upon which researchers decide whether behavioural measurements qualify as a personality trait or not. In behavioural ecology, personality is defined as between-individual differences in a behaviour in the average environment, that is, whether individuals show differences in the intercept of a behavioural reaction norm (given that the environmental predictors are centered or z-standardized; see manuscript I). If these differences are not entirely due to variation within individuals, that is, repeatability is higher than zero, they are considered to reflect personality. Zero repeatability of behavioural differences between individuals is very unlikely under most conditions and for most behaviours, because close to all of them carry a heritable component (Turkheimer, 2000) and (non-shared) environmental influences may contribute to their stability as well (Briley & Tucker-Drob, 2017). This means, that a distinction between a behavioural trait and a personality trait is rather artificial in behavioural

ecology, because about all behaviours fulfil such criterion. Since psychologists typically study more global personality dimensions (latent variables), this equivalence does not hold true for their discipline. We elaborate on this in manuscript I. Still, readers should note that heritability may be fairly low for specific behaviours (Dochtermann, Schwab, & Sih, 2015) and cases of non-significant stability of behaviour may occur (see e.g., van Dongen et al., 2010). Some studies used to infer the presence of personality from having a statistically significant estimate of stability (r or ICC / R), which, for most sample sizes, relies on effect sizes of moderate size or higher. It has, however, rightly been advised against any such threshold, since this would be arbitrary given that statistical significance is bound to sample size and effect sizes vary in meaning for different test-retest-intervals (e.g., Stamps & Groothuis, 2010). Most research questions in persoanlity psychology and behavioural ecology go beyond asking *if* there is any stability of behaviour within individuals, but asking *how much* behavioural stability individuals exhibit and how stability or change can be explained at various levels.

As we argue in manuscript II, it is useful to study ontogenetic personality development, or behavioural development in general, from a behaviour genetics perspective. In behaviour genetics, behavioural variation usually is partitioned into variation due to genetic effects, variation due to shared environmental effects (e.g., non-genetic influence of parents on siblings) and variation due to non-shared environmental effects (e.g., specific events an individual experiences). For humans we know that behavioural variation is almost entirely due to genetic and non-shared environmental effects, that is, shared environment plays a very minor role in personality development (Turkheimer, Pettersson, & Horn, 2014). If we build upon this knowledge and leave epigenetic effects and cases of gene-environment interplay aside, we can break down personality development mainly into stability due to genetic effects, stability due to non-shared environmental effects and instability / change due to non-shared environmental effects. Understanding within-individual variability in behaviour throughout lifetime is, at least in psychology, therefore often about identifying "environmental" events that are strong enough to push individuals away from their genetically influenced mean (Turkheimer et al., 2014). The crux here is, that gene-environment interplay, especially gene-environment correlation, is not uncommon, which makes it difficult to distinguish between developmental processes stimulated by the environment and developmental processes stimulated

by genetic effects. An example of a type of gene-environment correlation would be that individuals actively seek out specific environments based on their genetic predisposition ("niche-picking"; Scarr & McCartney, 1983).

In behavioural ecology, the study of behavioural stability is often motivated to understand why behaviour is not infinitely plastic (i.e., flexible), since this would allow individuals an optimal adjustment to differing contexts (Sih, Bell, & Johnson, 2004). If we consider the inverse of behavioural stability estimates as an upper end of flexibility, studying its lifetime development may add an important piece in the puzzle of understanding why behavioural adaptation to the environment is limited. Stamps and Groothuis (2010) give the example, that finding major personality changes around the time of maturity may generate hypotheses about the influence of gonadal hormones in behavioural stability or change. Other examples could be the finding of varying heritability of behaviour throughout ontogeny, which could help to identify in which developmental stages individuals show more or less behavioural responses to environmental variation. Flexibly changing a behavioural strategy to match the needs of a fluctuating environment is discussed to carry additional costs as having to sample for information (David & Dall, 2016). The ability to sample information, as in picking up cues from the environment, may however be linked to cognitive and motoric abilities or experience that first need to develop during ontogeny (especially if environmental cues are unreliable and afford trial-and-error learning; Penke et al., 2007). Theoretical considerations about the adaptive value of consistency in behaviour, an important area of research in behavioural ecology, also profit from empirical findings of temporal or cross-context correlations. For example, when weighing the adaptive value of flexible vs. consistent behaviour it could be useful to know how environmental conditions influence survival and reproduction, how such influence is actually tied to behavioural variation (see e.g., variation in aggressive and bold behaviour in relation to predation risk; Bell & Sih, 2007) and how conditions fluctuate throughout individual's ontogeny (Dingemanse et al., 2004). If we consider a scenario where predation risk of a population shows yearly fluctuation and we suspect that neophobic behaviour varies in its adaptive value due to this fluctuation, we could hypothesize that individuals show a reaction to predator fluctuation by varying their levels of neophobia. However, measuring behavioural reactions on a year by year basis may involve time-intervals that cover a large proportion of the average lifetime of individuals in some species.

Disentangling behavioural changes that occur due to fluctuation in context (predation risk) from temporal fluctuations (instability of a trait) may then be impossible, even if measurement error is thought to be low (Stamps & Groothuis, 2010). Here, we mentally revisit the crux described in the paragraph above: which proportion of temporal variation in behaviour is due to genetic effects and which due to environmental effects? Consider the example of genetically influenced differences in life history strategies, where individuals with a "live fast, die young" lifestyle show generally less flexibility in behaviour, they are always bold, whereas the "slow developers" show high flexibility. Now fluctuations in predation risk may affect individuals differently, based on their genetic background. Setting up experimental tests of behavioural flexibility adequately may hence be guided by combining a control for genetic effects (e.g., by using clones) and knowledge of lifetime behavioural development (e.g., by accounting for how flexible individuals are in behaviour across their life cycle).

In summary, both psychologists, working mainly on human personality, and behavioural ecologists, working mainly on animal personality, agree that studying causality in developmental processes of behaviour is a fairly complex endeavour due to potentially differing effects of temporal or spatial variation entangled with various forms of gene-environment interplay. Scarr and McCartney (1983), as well as Stamps and Groothuis (2010) provide many more examples of gene-environment interplay for the human and animal literature than I can do here. In manuscript II, where we reviewed studies of personality development in human and non-human primates, we unravel this topic also to a deeper extent. Likewise, in manuscript I, we discuss various aspects of temporal or contextual variation and co-variation in behaviour.

1.2 Evolutionary genetics of personality

Why are personality differences heritable, that is, passed on from generation to the next? In theory, natural selection could favour an invariant genome that codes for one single phenotype with optimal fitness, meaning that there is no variation which could be inherited. Yet, we know that all personality traits, at least those commonly described for humans, are heritable (Penke et al., 2007). By some mechanism genetic variation is hence maintained. Here, I will shortly summarize and discuss some of the mechanisms that have been proposed to maintain genetic variation in personality traits.

One possibility of finding heritable personality traits is that the variation within a trait is neutral to selection, that is, variation in such trait has (almost) no effect on reproductive success or survival in any environment. However, research in humans (Eaves et al., 1990; Roberts et al., 2007) and animals (Dingemanse & Réale, 2005; Smith & Blumstein, 2008) points into a different direction. That is, several personality traits are known to affect measures of evolutionary fitness. The case of selective neutrality is thus an implausible explanation (for a thorough discussion see Penke et al., 2007).

Having in mind that personality differences affect fitness and thus are evolutionary selected for, one could come to think that genetic variation is maintained because of mutations in the genome, since they are usually (mildly) harmful and work against selection, potentially leading to a mutation-selection balance. Such scenario is especially likely for traits that have a high mutational target size, that is, they rely on a large number of genetic loci which are potentially affected by a mutation. Throughout the last years it has become clear that personality traits in humans (Munafò & Flint, 2011; Sallis, Davey, & Munafò., 2018) and animals are largely polygenic (Bubac, Miller, & Coltman, 2020). There are however also aspects that contradict the plausibility of mutation-selection balance in personality. As Penke and colleagues (2007) argue, mutation-selection balance is a likely scenario for traits that reflect the overall functional integrity of the organism. Such „condition-dependant" traits have a high proportion of additive genetic variance because they are affected by large parts of the genome and thus have especially high mutational target size. In humans, personality traits are polygenic, but they are rather not related to the condition of the organism, are comparably unimportant in mate preferences, show thus rather low assortative mating, and no inbreeding or outbreeding effects. All in all, they are not very promising candidates for mutation-selection balance because they are not very closely linked to fitness and selection and don't have the very high mutational target size of condition-dependant traits (Penke et al., 2007).

A more likely explanation for finding personality variation in many species is balancing selection, which means that genetic diversity within populations is selected for (Fijarczyk & Babik, 2015). Balancing selection occurs due to variation in selection pressures that favour different phenotypes under different conditions, which leads to two or more different phenotypes with identical average fitness across environments (Penke et al., 2007). There are different types of balancing selection, the most suitable types

to explain genetic variation associated with personality are environmental heterogeneity and negative frequency-dependent selection (ibd.). Negative frequency-dependent selection occurs when the rare phenotype has a selective advantage, for example, because a rare behavioural strategy helps to avoid competition. Genetic variation associated with different behavioural strategies may thus be passed on to following generations. Environmental heterogeneity could be due to spatial or temporal variation in the environment. Spatial variation (e.g., variation of predation risk in different parts of the environment) combined with constrained ability to adjust behaviour to the local environment could lead to behavioural variation among individuals. Under temporal variation it is still likely that a single strategy maximizes evolutionary success across many generations, but variation in behaviour could occur when a single genotype produces a stochastic distribution of phenotypic expressions to match such fluctuations ("bet-hedging", Wolf & Weissing, 2010). Note, that although negative frequency-dependent selection and spatial heterogeneity can lead to genetic variation associated with behaviour, behavioural variation can also be maintained without variation on the genetic level in these scenarios. For example, when the same genotype allows a sufficient amount of phenotypic plasticity.

Another explanation of co-existing behavioural strategies is that they could be calibrated to other inherited traits of an individual, for example body size or strength. This could lead to adaptive combinations of physiological/morphological and behavioural traits. In behavioural ecology this is called state dependent behaviour, in evolutionary psychology facultative calibration. Following this theory, the heritability of such calibrated behaviour could be of a reactive kind, based on the heritability of the physiological or morphological traits of the individual (Tooby & Cosmides, 1990; Wolf & Weissing, 2010). In manuscript III we discuss facultative personality calibration in humans based on two empirical studies. In manuscript I we give additional examples of how evolutionary adaptive (co-)variation in behaviour may lead to variation on the genetic level.

1.3 Methods in personality research

In this section I will briefly introduce different ways of data collection and associated measurement concepts that are to a varying extent used by behavioural ecologists and psychologists in the study of

personality. Hereby, I roughly distinguish three different ways of data collection: questionnaire ratings, behavioural tests, and (non-experimental) behavioural observations.

Questionnaire ratings

The usage of questionnaires is the most common way of collecting data on personality differences in humans. The rationale is to ask an individual or their acquaintance to rate itself or the known person on a variety of different descriptions or adjectives (denoted as "items") in which humans are known to differ. Several items are typically aggregated to broader personality dimensions, in statistical terms, to latent variables. Individual assessment of personality is usually realised via pre-developed questionnaires that already include a balanced number of descriptions or adjectives that have been previously mapped to one of the different latent dimensions of an underlying structural personality model such as the Big Five or HEXACO (Lee & Ashton, 2004). Most questionnaires aim at capturing "typical" behaviour, thoughts or feelings of people (e.g., how impulsive they react on average), with thoughts and feelings being included as predispositions to behaviours. Questionnaire ratings of animals mirror this approach by asking human informants to rate animals on descriptive terms of the animal's behaviour (e.g., adjectives or behavioural descriptions; Koski, 2011; Uher, 2008). Raters are typically instructed to base their ratings on overall and average impressions, and not on estimated frequencies of specific behaviours or on particular situations (Freeman et al., 2013; Weiss et al., 2009). Importantly, these ratings should predict behaviour or "real-world outcomes" and show inter-observer agreement among independent raters (Gosling & Vazire, 2002). The most established questionnaire for animals is arguably the Humanoid Personality Questionnaire (HPQ), developed for non-human primates (King & Figueredo, 1997). Still, the applicability of this questionnaire needs to be reassessed for every new species or population, since research has shown non-negligible differences in the correlations between questionnaire items across species and populations (in animals see e.g., Weiss, 2017; in humans see e.g., Gurven et al., 2014; Lukaszewski et al., 2017).

Behavioural Tests / Experiments

Behavioural tests are the primary choice of data collection in behavioural ecology. They are designed to induce situations where individuals show behavioural differences in the wild. Prominent examples are open

field and novel object tests, where animals are measured in their propensity to approach and explore unfamiliarity, or playback experiments, where for instance the reaction to the recordings of predator calls is measured (Carter et al., 2013; Neumann, Agil, Widdig, & Engelhardt, 2013; Verbeek, Drent, Wiepkema, 1994). Typical examples of behaviours covered by such tests are boldness, exploration tendency, activity, or aggression. To estimate the repeatability of a shown behaviour, experiments need to be run several times, which may require statistical or experimental control for effects of habituation (Carter et al., 2013; Réale et al., 2007).

Behavioural tests that are designed to capture individual differences are also not uncommon in humans, although less prominent than the use of questionnaires. In psychology such tests are usually labelled Objective Personality Tests (OPTs), aiming to experimentally test expressions of personality traits that are captured in simulated miniature situations (Cattell, 1948; Ortner & Schmitt, 2014). This could, for example, be how accurate a person completes an intellectually non-demanding task (measuring conscientiousness; Koch et al., 2014) or helps others while working on an own assignment (measuring agreeableness; Poorthuis et al., 2014). Most modern OPTs are designed to measure a specific behaviour or trait on which humans are known to vary, rather than a multi-dimensional structure of personality as the Big Five (see Ortner & Schmitt, 2014).

Behavioural Observation

Behavioural observations of humans in non-experimental settings (or reported counts of a certain behaviour) are often collected in psychology, but seldom used to explore or structure a pattern of covariation among them or to assess their repeatability. This is likely because individual differences are typically studied within in the framework of latent personality traits, which are thought to comprise differences in behaviour, thoughts, and feelings altogether (see above). In personality psychology, observed behaviour is therefore often used as a validation criterion of a given personality trait, but not as its measurement per se (Back, Schmukle, & Egloff, 2009; Jackson et al., 2010). Additional ethical or practicable reasons may of course also hinder researchers to follow humans in their everyday life and to keep track of what they are doing. With the advent of electronic portable devices, assessment of everyday behaviour has

however become an increasingly used possibility in psychology. Under the label of "ambulatory assessment", different methods of electronically sampling behaviour are currently in use. One possibility is to ask people for self-reports of their daily behaviour via computer or smartphone ("experience sampling"). Other, more observational, methods involve recordings of movement, sound or video by portable electronic devices, or the content analysis of social media or text messages (see Wrzus & Mehl, 2015 for an overview). Traditionally, behavioural observations tended to play a bigger role in studies of behavioural development in children, where self-reports of personality are not a viable measurement option. Here, observed behaviour in experimental situations, recreation time, school or kindergarten may be the primary data source from which individual differences are quantified (see e.g., Reznick et al., 1986; Shoda, Mischel, & Wright, 1994). Behavioural observations of animals in their usual living conditions (either in captivity or the wild) are very common in some areas of behavioural ecology, especially in primatology (e.g., Brent et al. 2014; Ebenau et al., 2019; Tkaczynski et al., 2019; von Borell et al. 2016). Given sufficient amount of data on each individual's behaviour to ensure representativity, behavioural observations are especially valued for their high ecological validity (Koski, 2011).

1.4 Transition to the included manuscripts

I structured the introduction to capture different aspects of personality research on which the disciplines psychology and behavioural ecology tend to differ. These aspects will pop up repeatedly throughout all of the three manuscripts. Briefly, they are summarized into the following remarks.

1. Personality research in psychology and behavioural ecology use a different terminology and methodology. Manuscript I offers a translation and a comparison of empirical and conceptual work across disciplines.

2. The ontogenetic development of personality is most thoroughly studied in humans. A review and integration of findings from humans and non-human primates is offered in manuscript II.

3. Behavioural ecologists and psychologists share a common framework in evolutionary genetics, although research investments in this topic are somewhat differential. In manuscript III we empirically tested an evolutionary hypothesis of personality variation in humans, which can be conceptually transferred to studies in behavioural ecology.

Summary of manuscript I

In the first manuscript, we integrate the literature on personality research from psychology and behavioural ecology. First, we trace back how research goals are rooted in the historical traditions of these disciplines. Personality research in psychology aimed from the very beginning for a holistic description of individuals. By this means, the instruments designed to measure personality in humans capture several global traits, which are thought to cover most of the variation in behaviour we can observe among humans. Behavioural ecologists have been more interested in understanding why individuals differ in behaviour per se. Their studies did and do not try to deliver a holistic description of individuals, but rather test hypotheses under which circumstances variation among individuals in one or several behaviours can be explained. This historic divide is visible in the differences among disciplines regarding the definition of personality, related terminology, methodology, data structure and empirical results. We provide a translation of these differences throughout the manuscript and transfer conceptual ideas and findings from one "language" into the other. Hereby, we discuss evolutionary and proximal mechanisms of behavioural variation between individuals ("personality) and within individuals ("plasticity" and "predictability"). We can show that many empirical findings regarding the stability of behaviour and the influence of environmental circumstances on plastic responses are comparable across disciplines and species. We end with providing recommendations for future research in both disciplines that could lead to more comparable results.

Summary of Manuscript II

In this manuscript we review theories and findings on ontogenetic personality development in humans and non-human primates. Our focus is to explain how and to what extent genetic and environmental contributions interplay and shape such development. The main findings of our review are that non-human primates develop towards increasing rank-order stability and a pattern indicative of what has been described as a "mature" personality in humans, that is, becoming more agreeable, conscientious, and emotionally stable. Whereas environmental influences on behavioural variation among individuals act in humans especially around the time of adolescence and young adulthood, behavioural variation in non-

human primates seems to already be affected early in life. Among these early environmental influences are stress-related variation in the natural environment, parenting style or rearing conditions. Later in life, migration or maternity during young adulthood may also affect ontogenetic personality development. We show that the interplay among genotype and the environment is likely complex and ubiquitous in personality development and that causality in pathways that lead to behavioural differences are therefore difficult to pin down. We propose that one possible way to peek inside this "black box" is to conduct genetically informed longitudinal studies that are also adequately powered in the statistical sense. The latter has often been a problem in studies of non-human primates. As a possible solution we propose the identification of polygenic scores for behavioural differences in relatively large samples of a species, for example in breeding facilities, and then to apply this knowledge to the typically smaller populations in the wild or in other captive settings, such as zoos or sanctuaries.

Summary of Manuscript III

The third manuscript is concerned with the theory of facultative calibration. In this theory it is proposed that individual differences in personality may not be actively selected for, but are calibrated to differences in other traits of individuals, such as physical strength, physical attractiveness, or intelligence (Haysom et al., 2015; Lukaszewski & Roney, 2011; Sell et al., 2009; Tooby & Cosmides, 1990). These latter traits are thought to enhance the formidability of individuals, which we define broadly as the ability to inflict costs on others or to extract benefits from them. Previous literature showed convincing evidence for a relationship between self-rated measures of formidability (such as self-perceived physical attractiveness) and personality, but only mixed results or missing evidence for relationships between personality and other-rated or direct anthropometric measures of formidability. This means that the heritability of personality variation could be, contrary to the theory of facultative calibration, not reactive to heritability of other phenotypic traits, but solely related to how people perceive themselves in these traits. In the latter case the causal effect could go in the other direction, where individuals with certain personalities have a tendency to perceive themselves as more attractive or formidable, independent of their objective physique. We tested hypotheses of facultative personality calibration in two independent studies (N1 = 119 men and 124 women, N2= 165

men) and in an integrative way that included various personality outcomes of previous studies in the same samples (including Anger Proneness, Extraversion, Neuroticism, Narcissism, Shyness, Vengefulness, and Sociosexual Orientation), self-rated, and objective measures of formidability. Formidability measures were derived from assessments of physical strength and various anthropometric measures from full-body 3D scans and paired with measures of self-perceived and other-rated physical attractiveness (based on rotating morphometric 3D body models and facial photographs). We could replicate positive correlations with self-perceived attractiveness across outcomes, but these were not corroborated by more objective assessments of attractiveness. An effect of other-rated attractiveness was clearly not supported in our results for either sex, regardless of the personality outcome. Anthropometric measures and physical strength were also largely unrelated to personality, with the exception of Extraversion, Utility of Personal Aggression, and Sociosexual Orientation in men. The evidence across samples was however mixed. In one sample, domain-level Extraversion was related to strength and anthropometric measures in men, in the other sample only the Extraversion facets Activity and Assertiveness. For Sociosexual Orientation the results of our two samples varied more substantially, a positive association was only present in Study 2. Future studies need to clarify whether formidability, potentially an indicator of genetic quality for males, enhances their orientation and success in short-term mating.

Discussion

I would like to discuss the content of the three manuscripts alongside an example. Assume an individual that is, in comparison to others, likely to take risks in the face of uncertainty. This could be a human or animal of any species. A specific situation could be how this individual is willing to enter a potentially dangerous and harmful new environment. The common thread in all manuscripts is to describe and understand why an individual does so and other individuals do not.

If we observe between-individual differences in such "risk-taking" across repeated measurements, psychologists and behavioural ecologists will agree to face the phenomenon "personality". In manuscript I we discussed different uses of the term "personality". One use is to say, an individual is likely to take risks because of it's personality. This implies a more general disposition that influences the behaviour of an individual. As we argued, this belief is especially popular in psychology, but also behavioural ecologists may not refrain from placing risk-taking in a specific situation in a more general fearful-bold continuum on which individuals differ (Perals, Griffin, Bartomeus, & Sol, 2017).

Another use of the term "personality" is to describe the presence of between-individual variation in behaviour. A common definition in behavioural ecology (see the introduction to this thesis). This definition does not include a more general disposition affecting multiple behaviours. Following this definition, between-individual variation in risk-taking *is* personality, and not *because of* personality. The upside of such definition: it avoids a potentially simplified or incorrect assumption of a more common cause. We can see the merit, when we try to match risk-taking to one of the factors in popular personality models in humans. Does it belong to Fearfulness, a facet of Emotionality in the HEXACO Model? Or to Excitement Seeking, a facet of Extraversion in the Five Factor Model? But if we refrain from personality traits as global dispositions, how do we describe the tendencies of individuals to behave in a similar way across a variety of different situations? The problems we encounter in terminology are partly result of a causal manifold in behavioural variation. A manifold that likely entails both rather general / dispositional and more specific sources of variation. A disposition could be how responsive an individual generally is to stress or reward, based on underlying biological systems. A more specific source could be a salient experience an individual

Page 18

has made in the past. An individual should be more likely to engage in risky behaviour when the thrill activates the reward system, and/or the danger does not induce too much fear or stress response, and/or previous experiences in similar situations have been rewarding or at least not too aversive. As discussed in manuscript I and the introduction to this thesis, such combination of proximal biological mechanisms and the ontogeny of the individual are the classic building blocks of personality research in psychology. Behavioural ecologists, on the other hand, especially focus on evolutionary questions. Such questions could be: Why is it adaptive to be risk-taking? If we observe individual differences in risk-taking, how are they maintained throughout generations? Does taking risks and not taking risks have equal fitness outcomes? The question whether risk-taking is part of a more general disposition on which selection acts, i.e. being generally bold or fearful, is only one among many. There are thus many studies in behavioural ecology that do not use or test the concept of personality traits as latent variables, which are still predominant in psychology. Even if multiple behaviours are correlated in a study, behavioural ecologists do not necessarily assume a latent variable. Behavioural syndromes, which is how suites of correlated behaviour are referred to in behavioural ecology, may occur for other reasons than a common biological basis. An example would be that two behaviours are especially adaptive in combination but are not related on the genetic level (see manuscript I). This distinction among disciplines is one reason why "personality" remains, overall, an elusive term. Another reason is that personality is differentially attributed to variation at the between- or within-individual level, or both of them. This lead to substantial variation in terminology use also within behavioural ecology (Sánchez-Tójar, Moiron, & Niemelä, 2022). A differing use of terminology is also deeply embedded in different approaches to personality measurement. I briefly reviewed the most common approaches in the introduction. For the most part, behavioural ecologists favour a different approach (behavioural tests) than personality psychologists (questionnaires).

As outlined above, the study of individual differences in psychology looks back on a tradition of language-based assessment that (predominantly) tries to capture individual differences in self- or other-perceptions in a descriptive and comprehensive way. A questionnaire item like "I often take risks" will probably be one among many. Building a taxonomy of individual differences based on descriptors from the natural language leads to a data complexity that demands reduction. The long-time favoured theory that

various behavioural differences emerge from more basic differences in broad latent traits has been a good fit to handle such complexity, as it allows to summarize behavioural differences based on the correlational pattern between them. Without doubt, Big Five or HEXACO dimensions have proven themselves beyond being useful for data reduction. For example, they correlate with measurements of actual behaviour in a hypothesized way (Back et al., 2009) and predict various life outcomes (Roberts et al., 2007). Personality questionnaires in which one individual (rater) rates another individual (ratee) are essentially the same for humans and animals. Why don't behavioural ecologist use them? There are two main reasons. First, behavioural ecology has a strong focus on understanding how the behaviour of organisms is adapted to the environment (Simmons, 2014), including variation within individuals. For that matter, behavioural ecologists need to precisely manipulate or track different ecological conditions and sample behavioural measurements across these conditions. Retrospective ratings of human informants that reflect an overall impression of an individual across conditions or that are inaccurate in separating impressions caught in different conditions would be an ill-advised choice of data collection. If at all, questionnaires would need to reflect short-term ratings of animals. There is however, and this is the second aspect, belief that questionnaires are potentially subjective and biased towards the categorizations held possible by the researcher. Actual measurements of behaviour are, on the other hand, widely accepted in all research areas of behavioural ecology. Instead of questionnaires, standardized behavioural tests have therefore been proposed as an economical way to measure "personality" data (Réale et al., 2007).

Not to be mistaken, human personality research has not been blindfolded to the insight that different situations or environmental conditions likely elicit behavioural change. There are rather two differing approaches in psychology. For the scientific interest in hypotheses regarding cross-situational differences in broad traits, retrospective questionnaire ratings are typically used, while studies intended to explore the immediate interaction of personality and situations obviously track or manipulate the context in which a measurement takes place. For example, studies in children that employed behavioural tests (e.g., Fox et al., 2001; Kagan et al., 1988; Reznick et al., 1986) or observations in pre-classified situations (Shoda et al., 1994) provide very similar data to those in behavioural ecology studies. These studies are not a descendant of the sedimentation hypothesis that pursues a comprehensive description of individual

differences based on the analysis of natural language, but rather focus on (families of) basic traits in which early behavioural differences have been consistently observed in children. In adults, context-sensitive studies are still rather closely tied to the established latent factor models (mostly the Big Five) and the underlying questionnaires that are used ubiquitarily in psychology. Here, studies that take situational conditions into account typically stay within this framework by modifying existing questionnaires to meet the requirement of a short-term, i.e. situation-wise, reflection of behaviour (e.g., people are asked to rate their behaviour only for the past hour or within a specified situation, so-called *personality states*).

A follow-up question is whether different methods (i.e. questionnaires, behavioural tests, behavioural observations) would come to the same conclusion or even measure the same trait. In other words, how much are results influenced by the method used to obtain them? Quite a bit, as past research showed. For example, studies comparing personality self-reports and experimental approaches such as objective personality tests (OPTs) in humans found a lack of or only low convergence between them (e.g., Hofmann et al, 2005; Koch et al., 2014). As Koch and colleagues (2014) put it "[…] different types of methods (self-reports, implicit measures, and OPTs) seem to assess fundamentally different facets of the same attribute […]". Also in non-human primates the congruence between different measures has seen a lot of discussion and mixed results so far, where latent factors derived from behavioural observations and questionnaire ratings, with some exceptions, do not correlate as conclusively as expected (e.g., Garai et al., 2016; Iwanicki & Lehmann, 2015; Konečná et al., 2008; Uher & Asendorpf, 2008; Uher & Visalberghi, 2016; see Koski, 2011 for further references). A methodological bias influencing correlations among questionnaire ratings could occur due to semantic relatedness between behavioural descriptions. In such case, we would expect people or observers to rate themselves or an animal similarly on related adjectives and contrary on lexical antonyms to maintain semantic coherence (Kagan, 2003). Valuable analyses regarding this issue have been conducted by Weiss et al. (2012), who tested for anthropomorphic projections in personality ratings of orang-utans and chimpanzees. Briefly, they either subtracted an observer's average rating across all rated animals on an item from every single rating of an animal on that item ("M-Type analysis" reflecting only correlations based on differences between animals) or subtracted the average rating from all observers on an animal on an item from every single rating ("G-Type analysis" reflecting only differences based on

characteristics of the observers or item content). Latent factors derived from scores computed for the M-Type analysis showed overall high congruence with latent factors derived from the "standard" approach, where simply the mean of all observers' ratings for every animal on every item is entered into the analysis. The same comparison was done for scores computed for the G-Type analysis. Now, G-Type latent factors and standard latent factors should show low congruence if there are no effects of observers or item content. While the inter-factor correlations were overall not too worrying, the G-type factor structures, that is, which items loaded together on a latent factor, were clearly not arbitrary with respect to semantic content of the items. In conclusion, behavioural correlations based on questionnaire ratings likely reflect "true" effects of individual differences (see also Gosling & Vazire, 2002, on this matter), but semantic item content cannot be entirely ruled out to affect their magnitude or direction.

Of course, a mismatch between questionnaire ratings and other methods cannot be blamed on the questionnaires alone, behavioural observations and behavioural tests have their weaknesses as well. For example, they may suffer from selection effects, capturing only an incomplete fraction of the behavioural repertoire of a species or what has been covered by a questionnaire with a comprehensive descriptive aim (Bell, 2017). If so, it is not surprising when, for example, latent factors derived from questionnaires and latent factors derived from behavioural observations are incongruent. In such cases, it could arguably be better to validate latent questionnaire factors with directly measured behaviours than comparing factor structures with another. However, caution must be heeded with respect to the reliability of behavioural measures, which should match the usually high reliability of aggregated latent factors (Epstein, 1983). Possible solutions are to aggregate multiple behavioural measures which are hypothesized to correlate similarly with a latent factor (see e.g., Back et al., 2009; Jackson et al., 2010) or to aggregate multiple measurements within individuals. The importance of mixing and validating different ways of data collection has been stressed in the behavioural ecology literature as well. Perals and colleagues (2017), for example, tried to settle prior confusion regarding the biological meaning of metrics derived from open-field tests by contrasting them with behavioural measures from independent tests. A correlation between spatial exploration and latency to enter a novel space, they infer, would not reflect a common propensity of being explorative, but rather reflect a correlation between being explorative and bold. Whether behaviours are

common indicators of a more general trait or belong to functionally different dispositions may in turn affect hypotheses or questions about their co-occurrence, for example the question under which conditions it is adaptive to be both explorative and bold and under which not.

Now, let us consider an ontogenetic perspective on variation in risk-taking, following the lines of manuscript II. A central question in the study of "personality development" is whether individual differences in a trait are present from early on (i.e., infancy) and whether such differences remain stable or not. This question entails a link to the previous sections of the discussion, as, again, we have to consider what is meant by "a trait": a specific behaviour or a global disposition? In a slowly maturing species as humans, behavioural measurements of risk-taking will be different at different ages throughout ontogeny, because physical, cognitive or societal limitations in the expression of a given behaviour change with age. There will be presumably zero variability in how many times 2-year-old children have willingly jumped from a cliff into a lake or in speeding with cars. But there could be variation in biological underpinnings measured early in life that predict behaviour in adults. Following this logic, the "biological disposition" may not change but only the behavioural indications of such disposition. A thoroughly studied example of such a disposition in humans is called "behavioural inhibition". Inhibited children are cautious and shy in comparison to the more outgoing and fearless "uninhibited" children. Behavioural inhibition is an early precursor of aspects entailed in the adult personality dimensions introversion and neuroticism and has at least moderate stability throughout childhood and up to early adulthood (Fox et al 2005; Gest, 1997; Kagan et al. 1988). This stability can be partly explained by physiological differences among the inhibited and uninhibited, for example in the hypothalamic-pituitary-adrenal-axis (Kagan, Reznick, & Snidman, 1987). As indicated by the extended description "behavioural inhibition to the unfamiliar" (Kagan et al., 1984), this trait can be compared to the study of novelty avoidance in animals (Kagan, 1988). The reluctance to approach unfamiliar objects, spaces or people is amongst the most tested "personality traits" in animals, often also referred to as fearfulness/boldness (MacDonald, 1983). Although the disposition and associated physiology of being cautious and shy are relatively stable throughout ontogeny in humans, behavioural reactions or the salience of unfamiliar stimuli change with the age of the children. An example is that very young children (2-4 years old) who are shy tend to stay in close proximity to their parents when encountering an unfamiliar

situation, older children (above 6 years old) do less so and parental proximity is not tightly related to an inhibited disposition anymore (Kagan et al., 1987). Such pattern is termed heterotypic continuity (Putnam, 2011), that is, continuity (or stability/consistency) of individual differences in a broad trait, but a change in it's behavioural indicators. Other behavioural characteristics, as being quiet or avoidant of interaction in an unfamiliar social situation, tend to persist throughout older age groups. Such behaviours are more suitable to be studied in terms of homotypic continuity, that is, the stability of individual differences in a specific behaviour. These two types of continuity/stability are related to two different aspects under which behavioural development can be studied: mean-level and rank-order consistency. Mean-level consistency indicates how much the mean value of a behaviour expressed by a group of individuals changes throughout time, rank-order consistency the extent to which the rank order of scores for a given behaviour are preserved among individuals (Stamps & Groothuis, 2010). In the case of heterotypic continuity, behavioural indicators may show large changes in the mean and also in the rank-order among individuals when the meaning of variation among individuals in this behaviour changes with ontogeny. In case of homotypic continuity we may also expect large changes in the mean of the behaviour, but the rank-order among individuals in their behavioural expression should be relatively stable throughout time. A largely open question in psychology and behavioural ecology, at least in terms of empirical evidence, is whether the current utility of a behaviour (i.e., how adaptive a behaviour is) changes during ontogeny. Recall that a temporally fluctuating environment may maintain trait variation (see introduction). In addition, the utility of a specific trait expression could fluctuate with the age of the individual in a temporally stable environment. An example would be that explorative behaviour may have higher fitness costs at a young age, when the individual is for example less experienced in picking up environmental cues and physically immature, than at a later age. We can link the "learning / experience" part of this example with the idea of bayesian updating as a case of gene-environment interplay (see manuscript II) and the "physical maturation" part to the idea of facultative calibration of behaviour to physical formidability (see manuscript III). The idea of bayesian updating is that unexperienced ("naïve") individuals will vary in their phenotypic reaction to the environment based upon estimates of the value of environmental variables provided by their ancestors (via genes, epigenetic factors, and/or parental effects) and that these estimates are then updated based upon the

experiences of the individual across the lifetime (Stamps, Biro, Mitchell, & Saltz, 2018). The updated estimates of environmental variables allow individuals to change their phenotype in a way that is more likely to be adaptive under the actual conditions. There are various pathways of how such updating process could influence behavioural variation within and between individuals and it's associated utility. One example would be that initial variation in the expressed behaviour elicits variation in the reaction of conspecifics („evocative gene-environment correlation") which leads to updates of the expected social reaction to a behaviour and a) reinforcement of the behaviour if it turns out to have adaptive consequences or b) a change of strategy if the behaviour is rather associated with costs than benefits. Another example („active gene-environment correlation") would be that initial variation in the phenotype leads to variation in the environmental niches that are selected by individuals. Updating processes could then further increase variation in the experience with conditions of certain niches, which would increase the utility of the individual to stay within the initially selected niche or the exploration of novel niches. If we go back to the example of risk-taking we can imagine different foraging strategies that vary in the risk of survival. Risk-taking individuals may be more prone to explore the environment far from the initial home patch, generate information about these areas or how to behave in case of uncertainty and specialise in foraging in an explorative manner reinforced by increase in it's utility due to the updating process. Risk-averse individuals may be more prone to stay close to the home patch, generate information about this particular area of the environment and specialise in foraging in the home-patch. Updating could, for example, result in optimization of local foraging knowledge or information-related decrease in predation risk in the home patch and generate associated fitness increases. In manuscript III we showed that despite mixed results the best evidence for a relationship among physical formidability and personality was in the domain of extraversion, which is, in some of it's facets, related to risk-taking. In the end of the manuscript we discussed the causality of a formidability-extraversion relationship from an ontogenetic perspective. We can distinguish three possible pathways here. First, genetic differences among individuals lead to differences in formidability, based on which the expression of behaviour is calibrated. This is what advocates of the "facultative calibration theory" have in mind, that is, more formidable individuals can afford to show certain behaviour. Second, based on genetic and/or environmental differences and their interplay individuals differ

in behaviour which increases their formidability. For example, individuals who express an aggressive strategy towards conspecifics may exploit more resources, which leads to better physical condition. A third explanation would be mediation by another variable. An example thereof is a study that showed that boys with greater aggressive and antisocial tendencies at age 11 had greater increases in physical strength during the following six years of puberty, while not being consistently stronger than their peers at age 11. Here, the authors (Isen et al., 2015) proposed a joint hormonal mediation as explanation (see also manuscript III). Neither of these pathways is an exclusive explanation for the adaptive value of a formidability-behaviour relationship. In behavioural ecology such relationship is often discussed as state-behaviour feedback-loop (Sih et al., 2015), where physical condition and behaviour interact over time. The starting point could be either variation in state or variation in behaviour. Consider the example of risk-taking in foraging above. An initial difference in risk-taking can lead to increase in physical condition ("state") if the risky foraging is successful, in turn, this will give the individual enough resources to further engage in a risky foraging strategy and so on. An initial difference in state may also start of variation in risk-taking when only the individuals in good condition can afford such strategy and we are in the same feedback-loop as with initial differences in behaviour. The "mediation explanation" can easily be embedded in the literature on life-history theory and the pace-of-life syndrome in behavioural ecology, where covariation among physiological and behavioural traits (especially risk-taking) emerges because of differing investments in current or future fitness returns (Royauté et al., 2018; Wolf et al., 2007).

Final remarks

The common thread in my thesis is to integrate interdisciplinary perspectives on the question of why individuals differ in their behaviour. One of the main challenges, as I argued throughout, is a differential use of terminology and methodology among disciplines. In this last section I'll give a brief outlook on how such challenge can be met to jointly investigate "personality" from across disciplines in the future.

A first consideration is how study designs can comparably grasp research questions from both disciplines, despite the differential use of measurement instruments. If we opt for a common interest in the three major sources of variation shortly described above and in manuscript I, that is, personality, plasticity,

and predictability, we need repeated data of individuals both within and across varying "environmental" conditions. A suitable option for humans emerged during the last years with the applicability of experience sampling via mobile electronic devices. If people are asked about their behaviour at multiple instances and in conjunction with questions about the environmental circumstances they experience, we can calculate behavioural reaction norms or density distributions of behaviour. In manuscript I we discussed how personality psychology and behavioural ecology divide in that psychologists typically study "personality" in terms of broad domains or at least facets that include multiple questionnaire items. For example, Bleidorn (2009) asked the participants of her study to rate their behaviour, feelings, or thoughts during the last hour and the social role they occupied at multiple times a day and for several consecutive days, which allows for a separation of between-individual variation in the mean (personality), within-individual variation due to change in social role (plasticity) and within-individual variation while occupying the same social role (predictability). If the answers on questionnaire items are not aggregated to sum-scores reflecting higher-ordered aggregates (domains, facets), the main differences to studies in behavioural ecology are that participants rate their behaviour themselves on questionnaires and that the variance components have different names.

A second consideration therefore is whether we can find a consensus definition of personality (and associated concepts) that suits all researchers. Sánchez-Tójar and colleagues (2022), who discuss differential terminology use in behavioural ecology, propose a statistically oriented terminology that explicitly addresses whether behavioural variation is studied on the within- or between-individual level. This proposal accounts for the finding that some behavioural ecologists include limited plasticity (or generally variability) within individuals as part of a personality definition, whereas others define personality solely as between-individual variation. As outlined above and in manuscript I, differences in terminology use between behavioural ecology and psychology go beyond a within- vs. between-individual variation debate, because we additionally face a latent variable vs. no latent variable debate and a "personality as idiosyncratic description of an individual" vs. "description of variance-patterns" distinction. Also recall that the use of latent variables in psychology is partly due to the integration of (ratings of) feelings and thoughts, which are

not accounted for in animal studies. A simple set of definitions that could be used by both disciplines despite these differences is:

Personality: between-individual variation in the mean of a trait

Plasticity: within-individual variation in a trait due to variation in the environment

Predictability: within-individual variation in a trait that is not related to measured environmental variation

Psychologists and behavioural ecologists however use the term "trait" differently. Psychologists refer to the mean of a varying phenotype as "trait", whereas behavioural ecologists use the term more broadly, that is, not in referral to a central tendency in behaviour. Using "trait" as in the definitions above would mean that psychologists have to break-up with the common terminology of state-trait-theory (Steyer, Schmitt, & Eid, 1999) and adopt the term "trait" as a short equivalent of any phenotype on any hierarchical level irrespective of the phenotype's stability. In short: use "trait" as biologists do. This could however lead to confusion for practitioners, because of the widespread use of existing measurement inventories that carry a state-trait distinction. I therefore propose the following definitions:

Personality: between-individual variation in the mean of a behaviour or syndrome score

Plasticity: within-individual variation in a behaviour or syndrome score due to variation in the environment

Predictability: within-individual variation in a behaviour or syndrome score that is not related to measured environmental variation

The definition of personality is an extended variant of a common definition in behavioural ecology including the addition of variation in a syndrome. I chose the term syndrome, because it accounts for the findings and common practice in psychology to accumulate related phenotypes to a higher-ordered descriptive entity. The advantage of using syndrome instead of, for example, "personality trait" or "personality dimension" is that the other terms presumably lead to more confusion between disciplines. The term syndrome however has been used in behavioural ecology to describe the co-variation of multiple behaviours, is already established in psychology due to it's use in clinical psychology and does not imply a common cause assumption or latent variable, which is nowadays a strongly debated conceptualization also among

psychologists. The term "syndrome" is also not pre-occupied with expectations of the content of the related phenotypes, that is, it could be a syndrome of related behaviours (as in behavioural ecology), or related behaviours, feelings, and thoughts (as in psychology). Researchers could also freely decide how they want to summarize the covariation of related phenotypes belonging to a syndrome into a composite score. This could still be the value of a latent variable, but also other composite scores (e.g., summed values of questionnaire items, sum of z-standardised behavioural measurements).

The definition of plasticity captures the within-individual variation due to changes in the environment. Here I use the term "environment" in a broad sense and in relation to it's use in behaviour genetics, that is, "environment" describes all the organisms, objects, and conditions that surround an individual. Variation in the environment is also used irrespective of time, meaning that what psychologists often refer to as "situation" is encompassed in the use of "environment" as well. This reflects a broad definition of plasticity as the ability to adapt to environmental conditions (Frankenhuis & Nettle, 2020). Again, as for personality, we could chose to study plasticity on the level of a single behaviour or on the level of a syndrome. The latter would require some sort of composite score as a proxy of an individual's current state regarding the syndrome (see above). Hitherto, psychologists referred to such current state in a latent variable as "personality state", which can be very confusing for behavioural ecologists who think of personality as between-individual variation in the mean. I propose "syndrome state" as an alternative. As discussed above psychologists and behavioural ecologists use the term "trait" differently. To prevent misunderstanding, the more statistically oriented "mean" could replace the use of "trait" in psychology, that is, "syndrome states" vary around an individual's "syndrome mean".

Predictability would then be the residual variation within individuals in a (nearly) identical environment, for example, the variation among repeated measurements of the same test taken under nearly identical conditions. On a mere conceptual level one could argue that every variation within individuals is plasticity, because the conditions of repeated measurements are never completely identical. For example, a decrease in performance in a repeated test could be considered a plastic response to a diminishing energy budget of an individual. If we consider the statistical partitioning of variation, an

additional term of "predictability" can however be useful to describe the variation within individuals that are not accounted for by the environmental variables in the model (O`Dea et al., 2021).

References

Allport, F. H., & Allport, G. W. (1921). Personality Traits: Their Classification and Measurement. *The Journal of Abnormal Psychology and Social Psychology*, *16*(1), 6.

Back, M. D., Schmukle, S. C., & Egloff, B. (2009). Predicting Actual Behavior From the Explicit and Implicit Self-Concept of Personality. *Journal of Personality and Social Psychology*, *97*(3), 533–548. https://doi.org/10.1037/a0016229

Bateson, P., & Laland, K. N. (2013). Tinbergen's four questions: An appreciation and an update. *Trends in ecology & evolution*, *28*(12), 712–718.

Beekman, M., & Jordan, L. A. (2017). Does the field of animal personality provide any new insights for behavioral ecology? *Behavioral Ecology*, *28*(3), 617–623. https://doi.org/10.1093/beheco/arx022

Bell, A. M. (2017). There is no special sauce: A comment on Beekman and Jordan. *Behavioral Ecology*, *28*(3), 626–627. https://doi.org/10.1093/beheco/arx031

Bell, A. M., & Sih, A. (2007). Exposure to predation generates personality in threespined sticklebacks (Gasterosteus aculeatus). *Ecology Letters*, *10*(9), 828–834. https://doi.org/10.1111/j.1461-0248.2007.01081.x

Bleidorn, W. (2009). Linking personality states, current social roles and major life goals. *European Journal of Personality*, *23*(6), 509–530.

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*(5), 791–803.

Brent, L. J. N., Semple, S., MacLarnon, A., Ruiz-Lambides, A., Gonzalez-Martinez, J., & Platt, M. L. (2014). Personality Traits in Rhesus Macaques (Macaca mulatta) Are Heritable but Do Not Predict Reproductive Output. *International Journal of Primatology*, *35*(1), 188–209. https://doi.org/10.1007/s10764-013-9724-6

Briley, D. A., & Tucker-Drob, E. M. (2017). Comparing the Developmental Genetics of Cognition and Personality over the Life Span: Comparing Developmental Genetics. *Journal of Personality*, *85*(1), 51–64. https://doi.org/10.1111/jopy.12186

Bubac, C. M., Miller, J. M., & Coltman, D. W. (2020). The genetic basis of animal behavioural diversity in natural populations. *Molecular Ecology*, *29*(11), 1957–1971. https://doi.org/10.1111/mec.15461

Carter, A. J., Feeney, W. E., Marshall, H. H., Cowlishaw, G., & Heinsohn, R. (2013). Animal personality: What are behavioural ecologists measuring? *Biological Reviews*, *88*(2), 465–475.

Caspi, A., Roberts, B. W., & Shiner, R. L. (2005). Personality development: Stability and change. *Annual Review of Psychology*, *56*, 453–484. https://doi.org/10.1146/annurev.psych.55.090902.141913

Cattell, R. B. (1948). Primary Personality Factors in the Real of Objective Tests. *Journal of Personality*, *16*(4), 459–486. https://doi.org/10.1111/j.1467-6494.1948.tb02301.x

David, M., & Dall, S. R. X. (2016). Unravelling the Philosophies Underlying 'Animal Personality' Studies: A Brief Re-Appraisal of the Field. *Ethology*, *122*(1), 1–9. https://doi.org/10.1111/eth.12445

Dingemanse, N. J., Both, C., Drent, P. J., & Tinbergen, J. M. (2004). Fitness consequences of avian personalities in a fluctuating environment. *Proceedings of the Royal Society B: Biological Sciences*, *271*(1541), 847–852. https://doi.org/10.1098/rspb.2004.2680

Dingemanse, N. J., & Réale, D. (2005). Natural selection and animal personality. *Behaviour*, *142*(9–10), 1159–1184. https://doi.org/10.1163/156853905774539445

Dingemanse, N. J., & Wright, J. (2020). Criteria for acceptable studies of animal personality and behavioural syndromes. *Ethology*, *126*(9), 865–869.

Dochtermann, N. A., Schwab, T., & Sih, A. (2015). The contribution of additive genetic variation to personality variation: Heritability of personality. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1798), 20142201. https://doi.org/10.1098/rspb.2014.2201

Eaves, L. J., Martin, N. G., Heath, A. C., Hewitt, J. K., & Neale, M. C. (1990). Personality and reproductive fitness. *Behavior Genetics*, *20*(5), 563–568. https://doi.org/10.1007/BF01065872

Ebenau, A., von Borell, C., Penke, L., Ostner, J., & Schülke, O. (2019). Personality homophily affects male social bonding in wild Assamese macaques, Macaca assamensis. *Animal Behaviour*, *155*, 21–35. https://doi.org/10.1016/j.anbehav.2019.05.020

Epstein, S. (1983). Aggregation and beyond: Some basic issues on the prediction of behavior. *Journal of Personality*, *51*(3), 360–392. https://doi.org/10.1111/j.1467-6494.1983.tb00338.x

Fijarczyk, A., & Babik, W. (2015). Detecting balancing selection in genomes: Limits and prospects. *Molecular Ecology*, *24*(14), 3529–3545. https://doi.org/10.1111/mec.13226

Fox, N. A., Henderson, H. A., Marshall, P. J., Nichols, K. E., & Ghera, M. M. (2005). Behavioral Inhibition: Linking Biology and Behavior within a Developmental Framework. *Annual Review of Psychology*, *56*(1), 235–262. https://doi.org/10.1146/annurev.psych.55.090902.141532

Fox, N. A., Henderson, H. A., Rubin, K. H., Calkins, S. D., & Schmidt, L. A. (2001). Continuity and Discontinuity of Behavioral Inhibition and Exuberance: Psychophysiological and Behavioral Influences across the First Four Years of Life. *Child Development*, *72*(1), 1–21. https://doi.org/10.1111/1467-8624.00262

Frankenhuis, W. E., & Nettle, D. (2020). Integration of plasticity research across disciplines. *Current Opinion in Behavioral Sciences*, *36*, 157–162.

Freeman, H. D., Brosnan, S. F., Hopper, L. M., Lambeth, S. P., Schapiro, S. J., & Gosling, S. D. (2013). Developing a comprehensive and comparative questionnaire for measuring personality in chimpanzees using a simultaneous top-down/bottom-up design: Chimpanzee Personality. *American Journal of Primatology*, *75*(10), 1042–1053. https://doi.org/10.1002/ajp.22168

Garai, C., Weiss, A., Arnaud, C., & Furuichi, T. (2016). Personality in wild bonobos (*Pan paniscus*): Personality in Wild Bonobos. *American Journal of Primatology*, *78*(11), 1178–1189. https://doi.org/10.1002/ajp.22573

Gest, S. D. (1997). Behavioral inhibition: Stability and associations with adaptation from childhood to early adulthood. *Journal of Personality and Social Psychology*, *72*(2), 467–475. https://doi.org/10.1037/0022-3514.72.2.467

Giles, N., & Huntingford, F. A. (1984). Predation risk and inter-population variation in antipredator behaviour in the three-spined stickleback, Gasterosteus aculeatus L. *Animal behaviour*, *32*(1), 264–275.

Gosling, S. D., & Vazire, S. (2002). Are we barking up the right tree? Evaluating a comparative approach to personality. *Journal of Research in Personality*, *36*(6), 607–614. https://doi.org/10.1016/S0092-6566(02)00511-1

Gurven, M., Von Rueden, C., Massenkoff, M., Kaplan, H., & Lero Vie, M. (2013). How universal is the Big Five? Testing the five-factor model of personality variation among forager–farmers in the Bolivian Amazon. *Journal of personality and social psychology*, *104*(2), 354.

Haysom, H. J., Mitchem, D. G., Lee, A. J., Wright, M. J., Martin, N. G., Keller, M. C., & Zietsch, B. P. (2015). A test of the facultative calibration/reactive heritability model of extraversion. *Evolution and Human Behavior*, *36*(5), 414–419. https://doi.org/10.1016/j.evolhumbehav.2015.03.002

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A Meta-Analysis on the Correlation Between the Implicit Association Test and Explicit Self-Report Measures. *Personality and Social Psychology Bulletin*, *31*(10), 1369–1385. https://doi.org/10.1177/0146167205275613

Isen, J. D., McGue, M. K., & Iacono, W. G. (2015). Aggressive-Antisocial Boys Develop Into Physically Strong Young Men. *Psychological Science*, *26*(4), 444–455. https://doi.org/10.1177/0956797614567718

Iwanicki, S., & Lehmann, J. (2015). Behavioral and trait rating assessments of personality in common marmosets (Callithrix jacchus). *Journal of Comparative Psychology*, *129*(3), 205–217. https://doi.org/10.1037/a0039318

Jackson, J. J., Wood, D., Bogg, T., Walton, K. E., Harms, P. D., & Roberts, B. W. (2010). What do conscientious people do? Development and validation of the Behavioral Indicators of Conscientiousness (BIC). *Journal of Research in Personality*, *44*(4), 501–511. https://doi.org/10.1016/j.jrp.2010.06.005

Kagan, J. (2003). Biology, Context, and Developmental Inquiry. *Annual Review of Psychology*, *54*(1), 1–23. https://doi.org/10.1146/annurev.psych.54.101601.145240

Kagan, J., Reznick, J. S., Clarke, C., Snidman, N., & Garcia-Coll, C. (1984). Behavioral Inhibition to the Unfamiliar. *Child Development*, *55*(6), 2212. https://doi.org/10.2307/1129793

Kagan, J., Reznick, J. S., & Snidman, N. (1987). The Physiology and Psychology of Behavioral Inhibition in Children. *Child Development*, *58*(6), 1459. https://doi.org/10.2307/1130685

Kagan, J., Reznick, J. S., Snidman, N., Gibbons, J., & Johnson, M. O. (1988). Childhood Derivatives of Inhibition and Lack of Inhibition to the Unfamiliar. *Child Development*, *59*(6), 1580. https://doi.org/10.2307/1130672

King, J. E., & Figueredo, A. J. (1997). The Five-Factor Model plus Dominance in Chimpanzee Personality. *Journal of Research in Personality*, *31*(2), 257–271. https://doi.org/10.1006/jrpe.1997.2179

Koch, T., Ortner, T. M., Eid, M., Caspers, J., & Schmitt, M. (2014). Evaluating the Construct Validity of Objective Personality Tests Using a Multitrait-Multimethod-Multioccasion-(MTMM-MO)-Approach. *European Journal of Psychological Assessment*, *30*(3), 208–230. https://doi.org/10.1027/1015-5759/a000212

Konečná, M., Lhota, S., Weiss, A., Urbánek, T., Adamová, T., & Pluháček, J. (2008). Personality in free-ranging Hanuman langur (Semnopithecus entellus) males: Subjective ratings and recorded behavior. *Journal of Comparative Psychology*, *122*(4), 379–389. https://doi.org/10.1037/a0012625

Koski, S. E. (2011). How to Measure Animal Personality and Why Does It Matter? Integrating the Psychological and Biological Approaches to Animal Personality. *From Genes to Animal Behavior: Social Structures, Personalities, Communication by Color*, 115–136. https://doi.org/10.1007/978-4-431-53892-9_5

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate behavioral research*, *39*(2), 329–358.

Lukaszewski, A. W. (2013). Testing an Adaptationist Theory of Trait Covariation: Relative Bargaining Power as a Common Calibrator of an Interpersonal Syndrome. *European Journal of Personality*, *27*(4), 328–345. https://doi.org/10.1002/per.1908

Lukaszewski, A. W., Gurven, M., von Rueden, C. R., & Schmitt, D. P. (2017). What explains personality covariation? A test of the socioecological complexity hypothesis. *Social Psychological and Personality Science*, *8*(8), 943–952.

Lukaszewski, A. W., & Roney, J. R. (2011). The Origins of Extraversion: Joint Effects of Facultative Calibration and Genetic Polymorphism. *Personality and Social Psychology Bulletin*, *37*(3), 409–421. https://doi.org/10.1177/0146167210397209

MacDonald, K. (1983). Stability of individual differences in behavior in a litter of wolf cubs (Canis lupus). *Journal of Comparative Psychology*, *97*(2), 99.

Munafò, M. R., & Flint, J. (2011). Dissecting the genetic architecture of human personality. *Trends in Cognitive Sciences*, S1364661311001495. https://doi.org/10.1016/j.tics.2011.07.007

Nakagawa, S., & Schielzeth, H. (2010). Repeatability for Gaussian and non‐Gaussian data: A practical guide for biologists. *Biological Reviews*, *85*(4), 935–956.

Nesse, R. M. (2013). Tinbergen's four questions, organized: A response to Bateson and Laland. *Trends in Ecology & Evolution*, *28*(12), 681–682. https://doi.org/10.1016/j.tree.2013.10.008

Nettle, D., & Penke, L. (2010). Personality: Bridging the literatures from human psychology and behavioural ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1560), 4043–4050. https://doi.org/10.1098/rstb.2010.0061

Neumann, C., Agil, M., Widdig, A., & Engelhardt, A. (2013). Personality of Wild Male Crested Macaques (Macaca nigra). *PLoS ONE*, *8*(8), e69383. https://doi.org/10.1371/journal.pone.0069383

O'Dea, R. E., Noble, D. W. A., & Nakagawa, S. (2020). Unifying individual differences in personality, predictability and plasticity: A practical guide. *Methods in Ecology and Evolution*.

Orr, H. A. (2009). Fitness and its role in evolutionary genetics. *Nature Reviews Genetics*, *10*(8), 531–539. https://doi.org/10.1038/nrg2603

Ortner, T. M., & Schmitt, M. (2014). Advances and Continuing Challenges in Objective Personality Testing. *European Journal of Psychological Assessment*, *30*(3), 163–168. https://doi.org/10.1027/1015-5759/a000213

Penke, L., Denissen, J. J. A., & Miller, G. F. (2007). The evolutionary genetics of personality. *European Journal of Personality: Published for the European Association of Personality Psychology*, *21*(5), 549–587.

Perals, D., Griffin, A. S., Bartomeus, I., & Sol, D. (2017). Revisiting the open-field test: What does it really tell us about animal personality? *Animal Behaviour*, *123*, 69–79.

Poorthuis, A. M. G., Thomaes, S., Denissen, J. J. A., van Aken, M. A. G., & Orobio de Castro, B. (2014). Personality in Action: Can Brief Behavioral Personality Tests Predict Children's Academic and Social Adjustment Across the Transition to Secondary School? *European Journal of Psychological Assessment*, *30*(3), 169–177. https://doi.org/10.1027/1015-5759/a000186

Putnam, S. P. (2011). Stability and instability of childhood traits: Implications for personality development of animals. *Developmental Psychobiology*, *53*(6), 510–520. https://doi.org/10.1002/dev.20578

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

Reznick, J. S., Kagan, J., Snidman, N., Gersten, M., Baak, K., & Rosenberg, A. (1986). Inhibited and Uninhibited Children: A Follow-Up Study. *Child Development*, *57*(3), 660. https://doi.org/10.2307/1130344

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Important Life Outcomes. *Perspectives on Psychological Science*, *2*(4), 313–345. https://doi.org/10.1111/j.1745-6916.2007.00047.x

Royauté, R., Berdal, M. A., Garrison, C. R., & Dochtermann, N. A. (2018). Paceless life? A meta-analysis of the pace-of-life syndrome hypothesis. *Behavioral Ecology and Sociobiology*, *72*(3), 64. https://doi.org/10.1007/s00265-018-2472-z

Sallis, H., Davey Smith, G., & Munafò, M. R. (2018). Genetics of biologically based psychological differences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1744), 20170162. https://doi.org/10.1098/rstb.2017.0162

Sánchez-Tójar, A., Moiron, M., & Niemelä, P. T. (2022). Terminology use in animal personality research: A self-report questionnaire and a systematic review. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1968), 20212259. https://doi.org/10.1098/rspb.2021.2259

Scarr, S., & McCartney, K. (1983). How People Make Their Own Environments: A Theory of Genotype --> Environment Effects. *Child Development*, *54*(2), 424. https://doi.org/10.2307/1129703

Sell, A., Tooby, J., & Cosmides, L. (2009). Formidability and the logic of human anger. *Proceedings of the National Academy of Sciences*, *106*(35), 15073–15078. https://doi.org/10.1073/pnas.0904312106

Shoda, Y., Mischel, W., & Wright, J. C. (1994). Intraindividual stability in the organization and patterning of behavior: Incorporating psychological situations into the idiographic analysis of personality. *Journal of Personality and Social Psychology*, *67*(4), 674–687. https://doi.org/10.1037/0022-3514.67.4.674

Sih, A., Bell, A., & Johnson, J. C. (2004). Behavioral syndromes: An ecological and evolutionary overview. *Trends in Ecology and Evolution*, *19*(7), 372–378. https://doi.org/10.1016/j.tree.2004.04.009

Sih, A., Mathot, K. J., Moirón, M., Montiglio, P.-O., Wolf, M., & Dingemanse, N. J. (2015). Animal personality and state–behaviour feedbacks: A review and guide for empiricists. *Trends in ecology & evolution*, *30*(1), 50–60.

Simmons, L. W. (2014). 25 years of behavioral ecology. *Behavioral Ecology*, *25*(1), 1–3.

https://doi.org/10.1093/beheco/art108

Smith, B. R., & Blumstein, D. T. (2008). Fitness consequences of personality: A meta-analysis. *Behavioral Ecology*, *19*(2), 448–455. https://doi.org/10.1093/beheco/arm144

Stamps, J. A., Biro, P. A., Mitchell, D. J., & Saltz, J. B. (2018). Bayesian updating during development predicts genotypic differences in plasticity. *Evolution*, *72*(10), 2167–2180. https://doi.org/10.1111/evo.13585

Stamps, J., & Groothuis, T. G. G. (2010). The development of animal personality: Relevance, concepts and perspectives. *Biol Rev*, *85*. https://doi.org/10.1111/j.1469-185X.2009.00103.x

Steyer, R., Schmitt, M., & Eid, M. (1999). Latent state–trait theory and research in personality and individual differences. *European Journal of Personality*, *13*(5), 389–408.

Thompson, J. N. (1998). Rapid evolution as an ecological process. *Trends in Ecology & Evolution*, *13*(8), 329–332. https://doi.org/10.1016/S0169-5347(98)01378-0

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für tierpsychologie*, *20*(4), 410–433.

Tkaczynski, P. J., Ross, C., MacLarnon, A., Mouna, M., Majolo, B., & Lehmann, J. (2019). Measuring personality in the field: An in situ comparison of personality quantification methods in wild Barbary macaques (Macaca sylvanus). *Journal of Comparative Psychology*, *133*(3), 313–325.

https://doi.org/10.1037/com0000163

Tooby, J., & Cosmides, L. (1990). On the Universality of Human Nature and the Uniqueness of the Individual: The Role of Genetics and Adaptation. *Journal of Personality*, *58*(1), 17–67.

https://doi.org/10.1111/j.1467-6494.1990.tb00907.x

Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Current directions in psychological science*, *9*(5), 160–164.

Turkheimer, E., Pettersson, E., & Horn, E. E. (2014). A Phenotypic Null Hypothesis for the Genetics of Personality. *Annual Review of Psychology*, *65*(1), 515–540. https://doi.org/10.1146/annurev-psych-113011-143752

Uher, J. (2008). Comparative personality research: Methodological approaches. *European Journal of Personality*, *22*(5), 427–455. https://doi.org/10.1002/per.680

Uher, J. (2011). Individual behavioral phenotypes: An integrative meta□theoretical framework. Why "behavioral syndromes" are not analogs of "personality". *Developmental Psychobiology*, *53*(6), 521–548.

Uher, J., & Asendorpf, J. B. (2008). Personality assessment in the Great Apes: Comparing ecologically valid behavior measures, behavior ratings, and adjective ratings. *Journal of Research in Personality*, *42*(4), 821–838. https://doi.org/10.1016/j.jrp.2007.10.004

Uher, J., & Visalberghi, E. (2016). Observations versus assessments of personality: A five-method multi-species study reveals numerous biases in ratings and methodological limitations of standardised assessments. *Journal of Research in Personality*, *61*, 61–79. https://doi.org/10.1016/j.jrp.2016.02.003

van Dongen, W. F. D., Maldonado, K., Sabat, P., & Vásquez, R. A. (2010). Geographic variation in the repeatability of a personality trait. *Behavioral Ecology*, *21*(6), 1243–1250. https://doi.org/10.1093/beheco/arq145

Verbeek, M. E. M., Drent, P. J., & Wiepkema, P. R. (1994). Consistent individual differences in early exploratory behaviour of male great tits. *Animal Behaviour*, *48*(5), 1113–1121.

von Borell, C., Kulik, L., & Widdig, A. (2016). Growing into the self: The development of personality in rhesus macaques. *Animal Behaviour*, *122*, 183–195. https://doi.org/10.1016/j.anbehav.2016.10.013

Weiss, A. (2017). Exploring Factor Space (and Other Adventures) with the Hominoid Personality Questionnaire. In J. Vonk, A. Weiss, & S. A. Kuczaj (Hrsg.), *Personality in Nonhuman Animals* (S. 19–38). Springer International Publishing. https://doi.org/10.1007/978-3-319-59300-5_2

Weiss, A., Inoue-Murayama, M., Hong, K.-W., Inoue, E., Udono, T., Ochiai, T., Matsuzawa, T., Hirata, S., & King, J. E. (2009). Assessing chimpanzee personality and subjective well-being in Japan. *American Journal of Primatology*, *71*(4), 283–292. https://doi.org/10.1002/ajp.20649

Weiss, A., Inoue-Murayama, M., King, J. E., Adams, M. J., & Matsuzawa, T. (2012). All too human? Chimpanzee and orang-utan personalities are not anthropomorphic projections. *Animal Behaviour*, *83*(6), 1355–1365. https://doi.org/10.1016/j.anbehav.2012.02.024

Wolf, M., van Doorn, G. S., Leimar, O., & Weissing, F. J. (2007). Life-history trade-offs favour the evolution of animal personalities. *Nature*, *447*(7144), 581–584. https://doi.org/10.1038/nature05835

Wolf, M., & Weissing, F. J. (2010). An explanatory framework for adaptive personality differences.

Philosophical Transactions of the Royal Society B: Biological Sciences, *365*(1560), 3959–3968.

Wrzus, C., & Mehl, M. R. (2015). Lab And/Or Field? Measuring Personality Processes and Their Social

Consequences. *European Journal of Personality*, *29*(2), 250–271. https://doi.org/10.1002/per.1986

Appendix A: Manuscript I

Towards a future of interdisciplinary personality research – exchanging ideas between personality psychology and behavioural ecology

The study of behavioural variation based on differences among individuals is a common topic in both personality psychology and behavioural ecology. However, these disciplines however tended to develop independently from each other. As a consequence, researchers from each discipline may be confused by or unaware of terminology, statistical routines or empirical results from the other discipline. In this review, we aim to give readers from both disciplines a better understanding of how differences in the study of behavioural differences emerged, how different concepts and terminology relate to each other, and how advances in methodology and understanding of behavioural variation can be transferred across disciplines. Hereby, we will discuss evolutionary and proximal mechanisms of behavioural variation between individuals ("personality) and within individuals ("plasticity" and "predictability"). We highlight existing research topics on which both disciplines already converge and derive recommendations for future research.

1. Background

In both humans and non-human animals, individuals differ in how they behave in a given situation. If we observe such behavioural differences among individuals at multiple times, we may start to refer to such differences as "personality". There is quite some diversity in how "personality" is defined, both within and across disciplines that study this phenomenon. Despite terminological differences or subtleties, many scientists from different disciplines have used the term "personality" when studying variation in behaviour between individuals (for examples across animal taxa see, e.g., Bell, Hankison, & Laskoswki, 2009; Dingemanse & Wolf, 2010; Groothuis & Carere, 2005; Kralj-Fišer & Schuett, 2014; for humans, e.g., Caspi, Roberts, & Shiner, 2005; Funder, 1991; Goldberg, 1990). In this review we aim to integrate knowledge about personality from psychologists, mostly studying humans, and behavioural ecologists, mostly studying non-human animals (from here on referred to as "animals"). Frankly, this is not a novel endeavor (see e.g., Carter et al., 2013; Gosling, 2001; Koski, 2011; Nettle & Penke, 2010; Uher, 2008; Uher, 2011; Međedović, 2018, Frankenhuis & Tiokhin, 2018), but we believe that communication between disciplines is still marked by misunderstandings and should be a constant process, to keep the dialogue up to date. We will focus on three main aspects. First, we try to promote understanding of why researchers from different disciplines approach personality from different angles by providing a historical outline of personality research in psychology and behavioural ecology. Second, we translate concepts and jargon in the study of behavioural variation among disciplines in order to understand each other better in the future. Third, we try to integrate knowledge from both disciplines, which hopefully fosters cross-fertilisation in future studies.

2. Research goals and history lessons

In the early 1920's the study of personality gained in interest as a complement to the study of intelligence in humans. Intelligence was used as a label to describe individual differences in the ability of solving problems, e.g., of the school curriculum (Thorndike, 1924). Tests of intelligence were designed to provide quantitative information about performance differences in such ability. In distinction to intelligence, Allport and Allport (1921) described personality as an "elusive" term capturing individual differences that are only vaguely, if at

all, related to solving problems in the sense of an ability on which individuals fare better or worse. Rather, personality was used to describe "qualitative" differences among individuals, where "personalities of divers sorts succeed equally well in the general adaptation to situations of practical life" (Allport & Allport, 1921). Allport and Allport (1921) compared personality to a face that has no duplicates, an individual's unique mixture of varying degrees in diverse traits. They later go on to "find all the dynamic energy of the evolutionary struggle for existence condensed into one human personality", which emerges as a reaction of inherited behavioural tendencies to the (social) environment. Given that they identified personality as a mixture of traits that are I) partly inherited and II) involved in the adaptation to "situations of practical life", they could have started to sort out and test hypotheses about, say, which expression of a personality trait is favoured in which situation, why different mixtures of trait expressions fare equally well in life, as they propose, and so on, essentially asking questions about the evolution of "personality traits". Instead they aspired a holistic description of an individual's personality, that is, getting to know people really well by measuring them in a variety of different traits. Here, we may see the legacy of psychologists in their profession of providing counselling, self-understanding and therapy for their clients. In their inventory for personality assessment from 1921, Allport and Allport include, for example, differences in being introverted vs. extraverted, a distinction they took from the writings of Jung (1915). Their chief aims, as they write, were: "first, a picture of the individual personality and its checking up by the questionnaire or other information; second, the discovery of striking or unusual personalities and strongly contrasting personalities; and to find out what general types if any the various combinations of traits may reveal". These aims have clearly influenced the study of individual differences in psychology ever since and also reflect its link to practical usage. The Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013), used by clinicians, includes chapters on personality disorders, which can be perceived as extreme ends of "normal" personality patterns (Oldham & Morris, 1995), subsumed into diagnostic categories or "syndromes" (Shedler et al., 2010). Current developments in the conceptualization of psychiatric diagnoses, as represented by the Hierarchical Taxonomy of Psychopathology (HiTOP; Kotov et al., 2021) and Research Domain Criteria framework (RDOC; Cuthbert & Insel, 2013), advocate for a spectrum between "normal" and "pathological" personality. Baumgarten

(1928), a pioneer in human personality research, approached the topic via the assessment of occupational aptitude, trying to find a good fit between the characteristics of potential candidates and requirements of a job. Early work in psychological science followed Allport's and Allport's aims by trying to build a descriptive taxonomy of human personality. To reach this goal, they followed a logic which has become known as the lexical approach (or, alternatively, the sedimentation hypothesis). The idea was to build a scientific taxonomy based on natural language, reasoning that individual differences that are the most significant in daily transactions of people will eventually become encoded in their language (Goldberg, 1982; Klages, 1926). In extensive work, Baumgarten (1933; for the German language) and Allport and Odbert (1936; for the English language) tried to extract all terms of the dictionary that can be used to characterize a person (and its differences compared to others) and to summarize related terms to personality dimensions based on self- and other-ratings of individuals on these terms (see John, Goldberg, & Angleitner, 1984). In direct reference to the statistical method used, these dimensions are also called factors, as they emerged from factor analysis carried out to summarize the shared variance between terms in a correlation matrix (Cattell, 1946). The motivation behind summarizing terms was described by Cattell (1946) as an "unhappy necessity" to reduce the extensive list of verbally defined traits to a list compatible with rating them and to provide a parsimonious description of personality based on "major" traits (Cattell, 1943). The very popular five factor personality taxonomy is a modern descendant of this approach, which describes an individual's personality in five broad dimensions: Neuroticism (or, reverse coded, Emotional Stability), Extraversion (or Gregariousness/Surgency), Openness to Experience (or Intellect/Culture), Agreeableness, and Conscientiousness (Digman, 1990; Soto, Kronauer, & Liang, 2016; see Block, 1995 for a critical view). These five factors (or related taxonomies, such as the six-factor HEXACO structure, which adds Honesty-Humility; Lee & Ashton, 2004) are designed to measure "global traits" of people, i.e., overarching patterns rather than single or only narrow sets of behaviour (Funder, 1991). Given that the data underlying these taxonomies are often based on self-reports, they do not only encompass behaviour (which can be observed by others as well), but also feelings and thoughts (Jackson et al., 2010). Again, this also reflects popular models of psychiatric disorders, describing how thoughts or beliefs are connected to feelings and behaviour of an individual and are therefore viewed as part of a syndrome (Beck & Haigh, 2014; Ellis, 1991).

Beginning in the 1970's, comparative psychologists and ethologists worked on the establishment of personality taxonomies for animals. They subjected observed behaviour or behavioural ratings of non-human primates to factor analyses and obtained personality dimensions similar to those of humans (Buirski et al., 1973; Chamove, Eysenck, & Harlow, 1972, Stevenson-Hinde & Zunz, 1978). These findings motivated to study phylogenetic links between the personality structure of non-human and human primates (Weiss et al, 2011; Adams et al., 2015) and to explore personality structures in other species. Early examples being studies of cats (Feaver, Mendl, & Bateson, 1986), dogs (Goddard & Beilharz, 1984), wolves (MacDonald, 1983), rodents (Negrão & Schmidek, 1987; Royce & Poley, 1975), or various livestock (see Gosling, 2001 for a detailed list of studies). These studies fall within a time where the discipline of ethology, dedicated to the study of animal behaviour, was at a turning point, as researchers became increasingly interested in evolutionary and ecological aspects of behavioural variation, which led to the foundation of the discipline of behavioural ecology (Simmons, 2014). Clark and Ehlinger (1987, p.11) noted that studies of humans and non-human primates often addressed individual differences, but "their motivations differ from behavioural ecologists' interest in cost-benefit explications of behaviour […], the description and proximate causes of variation are carefully investigated, but evolutionary questions are lacking." As we argued above, such questions could have followed from the early thesis by Allport and Allport in 1921. The idea to study individual differences in non-human animal behaviour from an ultimate perspective, that is, to explore how selection acts on it, was instead proposed to the ethology and behavioural ecology literature.

Slater (1981) argued that behavioural variation could emerge because different behavioural strategies are favoured depending on the dynamics of competition among individuals (which is linked to evolutionary game theory, e.g. the hawks-doves game, Smith & Price, 1973) or through variation in selection pressures in different or fluctuating environments. Empirical studies from the following years supported this view. For example, Giles and Huntingford (1984) showed that three-spined sticklebacks differ in the population mean of fright response behaviour between populations that occupy different habitats with differences in

predation risk (see also Huntingford & Giles, 1987). Specifically, sticklebacks from high risk habitats show higher fright responses than those from low risk habitats. This evidence was seen as adaptive response that could be evolutionarily selected for by means of natural selection. Interestingly, they used principal component and factor analyses to summarize their behavioural measurements of fright responses into compound scores, because they wanted to simplify their data structure and facilitate statistical analysis and interpretation. We know this rationale from earlier studies in psychology (Cattell, 1943; 1946; see above). Wilson (1998) expanded the view of adaptive behavioural variation to within-population differences between bluegill and pumpkinseed sunfish individuals. Some individuals take more risk to feed in the presence of a potential predator, while others tend to hide and not feed in such situation. The "shy" individuals are arguably less likely to be killed by predators but have a limited access to food compared to the "bold" individuals, who however are at higher risk of predation. If the density of fish in a lake leads to high intra-specific competition, such mixture of behavioural strategies among individuals can adaptively evolve through a process called negative density-dependence. This again dovetails with Allport's and Allport's (1921) view that differences in personality cannot (always) be quantified in terms of better or worse, but as different strategies or mixtures which may fare equally well under certain conditions. More recently published examples underpinning Slater's (1981) case of fluctuating environments show how yearly variation in food abundance favours different ends of a behavioural spectrum in shyness vs. boldness regarding survival and reproduction, also leading to the maintenance of between-individual variation (Dingemanse, Both, Drent, & Tinbergen, 2004; Boon, Réale, & Boutin, 2007). Wilson (1998), however, argued against summarizing multiple traits by factor analyses into more general domains. He argued that selection pressures that favour individual differences in one context may differ from selection pressures in another context and should thus be studied separately for each "adaptive problem" of a population. Findings of uncorrelated between-individual variation in threatening versus nonthreatening contexts (i.e., contexts that pose different adaptive problems) was seen as support for this rationale (Coleman & Wilson, 1998; Sinn & Moltschaniwskyj, 2005).

Psychologists know this argument of context-dependency, it has been widely discussed in the literature as the "person-situation debate". This debate revolved around the question of whether individuals show stable and predictable behaviour across different contexts, that is, whether behaviour is due to an individual's disposition, or whether it is rather the specific environment or situation that shapes individual behaviour. Studies on cross-situational consistency showed that correlations among behavioural measurements at two measurement occasions or across two contexts typically range between 0.20 and 0.40 (Furr & Funder, 2004; see Fleeson & Noftle, 2008 for a short review). This magnitude of repeatability or cross-context stability has been considered small and taken as an argument for a situationist's perspective. The impact of situations on behaviour did however not produce higher numbers. Linear relationships between situational gradients and behaviour similarly produce correlations that do not exceed 0.40 (Funder & Ozer, 1983). Behavioural variation observed in a population is thus partly due to repeatable variation among individuals and partly due to flexible variation within individuals. Fleeson (2001) proposed to study "personality" as the whole density distribution of behaviour rather than focussing only on one parameter of the distribution, that is, not focussing on the mean or variation of behaviour in isolation. Based on repeated measurements of behavioural self-reports, he could show that the correlations between two randomly selected self-reports of behaviours (but accounting for auto-correlation) rarely exceeded 0.30 (see above). If, however, all self-reports (= measurements) were randomly split into two halves and the means of these halves were calculated, they were highly correlated (approx. 0.90 across behaviours).

The finding of both large behavioural variation within individuals and highly stable average tendencies in behaviour has been considered a synthesis resolution to the person-situation debate, integrating both the presence of flexibility and personality in human behaviour (Fleeson & Noftle, 2008). Behavioural ecologists have emphasized this as well. Dingemanse et al. (2010) conceptualized a conjunct study of between-individual variation in average behaviour and within-individual behavioural variation due to environmental variation in the *behavioural reaction norm* framework. A behavioural reaction norm can be visualized as a regression line with a certain elevation and slope. The elevation (or intercept) indicates the mean behavioural response of an individual exhibited in the average environment, provided that all

(environmental) predictor variables have been centred or standardised to a mean of zero. The slope indicates the behavioural variation of an individual along a varying environmental gradient. Based on this framework, behavioural ecologists typically define the variation of average behaviour between individuals, that is, variation in the intercepts of individuals at a mean-centered environmental gradient, as variation in *personality,* and the variation within individuals due to environmental variation as *plasticity* (Dingemanse et al. 2010; Dingemanse & Wright, 2020; but see Sánchez-Tójar, Moiron, & Niemelä, 2022, who have criticized ambiguous terminology within behavioural ecology). Variation that occurs within individuals but in the same environmental condition is referred to as *residual within-individual variation* (Westneat et al., 2015) or, reversely coded, as *predictability* (O'Dea, Noble, & Nakagawa, 2021).

It has been repeatedly pointed out that using the definition of personality as the average behaviour of an individual requires multiple measurements of behaviour, unless the behaviour is highly repeatable, that is, does not or hardly vary within individuals. In such case, a single measurement of a behaviour would be a proxy of the individual mean of a behaviour. This is highly unlikely, because animals (including human animals) usually show considerable within-individual variation (Brommer, 2013; Dingemanse & Wright, 2020; Niemelä & Dingemanse, 2018). Behavioural ecologists often quantify the amount of variation due to differences between individuals with the *repeatability* coefficient, which is estimated as the amount of between-individual variance divided by total phenotypic variance (also called *intra-class correlation ICC*; Nakagawa & Schielzeth, 2010). If there was no variation within individuals, the repeatability coefficient would equal 1. The meta-analytical average repeatability of variation in animal behaviour across measurements is about 0.4 (Bell et al., 2009; Holtmann, Lagisz, & Nakagawa, 2017). This is of the same magnitude as the correlation of two behavioural self-reports in humans (see above) and indicates that the larger part of the total behavioural variation (60%) happens within individuals. Thus, although psychologists and behavioural ecologists have focused on fairly different research goals, they have come to very similar results regarding behavioural variation so far. Differences between disciplines therefore seem to rather stem from different aims, terminology and methodology than from differences between human and non-human animals. We will elaborate on this in the following section.

3. Behavioural variation between and within individuals – a translation of concepts and terminology

Psychological studies usually use self-reports to collect data on personality differences between individuals. If studies include measurements of actual behaviour, these are often not used as a measurements of personality per se. This is because personality traits are traditionally conceptualized as latent variables, which are attributes that cannot be directly observed but cause individual differences in behaviours, thoughts or feelings that are linked to them (Borsboom, Mellenbergh, van Heerden, 2003). Following this logic, measures of behaviour have often been used to validate personality traits, operationalized as a test of whether behavioural measurements are associated with personality traits in a hypothesized way (Back, Schmukle, & Egloff, 2009; Jackson et al, 2010). Whether latent variables accurately describe personality variation is debated in the behavioural sciences and we will turn to this topic below. For now we mention it here, because interdisciplinary differences in terminology and measurement follow from this perspective.

Oftentimes psychologists ask people to rate their personality in cross-sectional study designs that do not include several measurement occasions. If so, they are asked to indicate how they usually behave, feel, or think, that is, they are prompted to mentally summarize across time and situations. This allows to obtain estimates of average behaviour, feelings or thoughts without taking repeated measurements. These estimates serve as proxies of a cross-situationally consistent personality trait. The term *trait* directly refers to the average and consistent part of an individual's personality. Deviations from this average tendency are labelled personality *states*. Personality states are also often measured with questionnaires that prompt people to report only short-term or situation-wise occurrences of behaviour, feelings, or thoughts (e.g., people are asked to rate their behaviour only in the current situation). A personality state is thought to depend on the underlying trait (the cross-situationally consistent disposition), the situation, and the interaction between person and situation (Steyer, Schmitt, & Eid, 1999).

In behavioural ecology, a *trait*, for example a behaviour, may also vary within individuals, that is, the term does not only represent a consistent or average part of a distribution. There is hence no other term to describe a single or current measurement of behaviour which would compare to a personality *state*. Rather, state is used to describe an organism's current condition in general, which may encompass numerous variables, for instance size or energy reserves, but also aspects of the environment (Dall, Housten, McNamara, 2004). As outlined above, personality variation is defined as variation in individual's average behaviour across repeated measurements. If we, for now, equate personality differences in humans as differences in their behaviour (and not their thinking or feeling, which psychologist usually include), then personality as defined in behavioural ecology would translate to the personality trait level in psychology, that is, the cross-situationally consistent tendency to express a behaviour. What psychologists call a personality state could be any behavioural measurement taken by a behavioural ecologist, be it in the same environment at another time or in a different environment.

Scientists from both disciplines typically associate personality with a certain stability in behaviour. Behavioural ecologists quantify this stability with the repeatability coefficient, which is estimated as the amount of between-individual variance divided by total phenotypic variance (see above). Behavioural ecologists sometimes speak of *personality traits* when they mean behaviours in which individuals show repeatable between-individual differences (a repeatability coefficient larger than zero). We argue against such use. Given that we will find heritable variation in every behaviour provided enough measurements (see e.g., Turkheimer, 2000), every behaviour will be to some extent repeatable and this would replace one term (behaviour) with another (personality trait). This would not be true for most psychologists who study personality on a higher-ordered hierarchical level, that is, a personality trait represented by a latent variable based upon multiple indicators does not equate to the term behaviour.

Because in behavioural ecology, personality is solely used to describe a pattern of behavioural variation between individuals and a personality trait is redundant to a behavioural trait, a single individual does not

vary in its personality. Personality is not an individual-centred concept, the term is not used to describe an individual' s unique mixture of behavioural dispositions. This is a clear distinction from how the term personality is used in psychology. Whereas behavioural ecologists usually study variation of actual measurements of behaviour, psychologists usually study variation of latent factors (which are estimated from multiple manifest indicators, e.g., questionnaire items). In psychology, individuals therefore vary in their personality in the sense of variation in measurements of latent personality factors (see Table 1 for an overview of definitions). A special additional case would be the study of how individuals change in their rank-order of differences in the long term. For such studies the term personality development is used.

As we outlined above, behavioural variation within individuals may occur due to changes in the environment, because it could be adaptive to behave differently in different environments. In the behavioural reaction norm framework of behavioural ecologists, such behavioural variation is termed plasticity. In personality psychology the term is less common, but occasionally used, for example in the (revised) sociogenomic model of personality by Roberts (2018). In this model Roberts distinguishes between fluctuating, elastic and pliable plasticity representing moment-to-moment, extended but not permanent, and permanent changes in behaviour, feelings, or thoughts (i.e. latent personality factor scores). Personality psychologists often refer to such plastic deviations from the average level an individual scores on a personality factor (e.g., extraversion) as personality states (see above). When behavioural ecologists study behavioural plasticity, that is, assess how individuals adapt their behaviour to changes in the environment, this roughly equates to what personality psychologists call a study of variation in personality states (Bleidorn, 2009; Geiser et al., 2015). Within the behavioural reaction norm framework, behavioural ecologists call all the variation that cannot be explained by the environmental predictors in the statistical model *residual within-individual variation* (see above).

In Figure 1, residual within-individual variation occurs within each of the five positions of the environmental gradient. Such variation could occur because of some sort of "error" in the measurement procedure, for

example, an observer making a mistake or a participant of a study misinterpreting a questionnaire item. In psychological measurement theory, residual variation is therefore often referred to as *measurement error*, although the hidden influence of other variables is acknowledged as well. The latter, that is, meaningful but hidden sources of variation, are an important topic in behavioural ecology. If the residual within-individual variance is not homogenous alongside the reaction norm, as is the case for Individuals 1 and 2 in Figure 1, this could be an indicator of multidimensional reaction norms, that is, behavioural variation aligned to several different environmental gradients (see Westneat, Wright, & Dingemanse, 2015). Residual within-individual variation (or low "predictability") could however also occur due to variation in *internal states,* for example hormone levels, which are known to be highly variable within individuals and act as mediators of short-term adjustment in behaviour (Holtmann et al., 2017).

For humans we know that overall within-individual variability (i.e. variation in "personality states", reflecting potentially both plasticity and residual within-individual variation) is highly stable. It has been shown that if repeated personality measurements are randomly split into two halves, not only the mean, but also the standard deviations of people's distributions of measurement values are highly correlated within individuals (Fleeson, 2001). Comparably, behavioural ecologists studying behavioural variation in the reaction norm framework found that the slopes of the reaction norm (i.e. plasticity) are temporally consistent ("repeatable") as well, although the magnitude of consistency varies with the environmental gradient under study (Araya-Ajoy & Dingemanse, 2017; Mitchell & Biro, 2017). While there is solid evidence that individuals also differ in their predictability across species and behaviours (Mitchell, Beckmann, & Biro, 2021), the temporal consistency of such differences in predictability has not been studied conclusively yet. A recent meta-analysis showed that the correlation of predictability estimates across different behaviours is rather small (Mitchell et al., 2021), that is, non-human animals do not appear to be generally predictable or unpredictable.

Results for humans are similar. Deviations from the average value in a personality factor (i.e. the "trait-level" in personality psychology) are also correlated in humans. Such correlations are usually higher when measurements are taken in functionally equivalent situations than in functionally different situations (Horstmann et al., 2021). From these results we can infer that humans tend to show consistent behavioural plasticity in response to a known situational affordance and to a lesser but still detectable amount a consistent behavioural response in situations where we do not know or do not have measured the reason for their behavioural adaptation. In the words of behavioural ecologists, humans show repeatability in behavioural plasticity and predictability. Alike non-human animals, humans do not have a general tendency of being plastic or predictable across behaviours or personality factors (Horstmann et al., 2021).

What Fleeson (2001) showed for humans was that the standard deviation of personality states is stable when they are aggregated across time and different situations. This, however, conflates two sources of within-individual variation that are studied separately by behavioural ecologists, plasticity and predictability. Although the magnitudes of plasticity and predictability may be correlated (O'Dea et al., 2021), that is, individuals could be generally flexible (high plasticity and low predictability) or generally inflexible (low plasticity and high predictability), this does not have to be true. In an experimental study, Mitchell and Biro (2017) showed such correlation for variation in activity rates in response to temperature change, but not (i.e., not statistically significant) for food deprivation, in zebrafish. Transferring this back to research in humans, there could be different reasons for finding a stable standard deviation of personality states for every personality factor in question. Whether an individual's standard deviation is stable due to stability in the response to measurable environmental predictors (referring to stability in plasticity) or not (which would mean high stability in residual within-individual variance / predictability) remains to be shown. Behavioural ecologists have also shown that predictability may change with a change in an environment gradient (e.g., residual variation decreases within individuals in reaction to environmental change; Mitchell & Biro, 2017). In other words, the density distribution and associated standard deviation of behavioural measurements may differ alongside the behavioural reaction norm (as for individuals 1 and 2 in Figure 1) and also differently in reaction norms of different behaviours.

The idea of an interplay between an individual's behavioural variability and the affordance of a situation has been, for example, conceptualized in the Nonlinear Interaction of Person and Situation (NIPS) model by Blum and colleagues (2018) in the psychology literature. This model rests on two premises: First, the magnitude of between- and within-individual variation in a behaviour is thought to vary with situational affordance, that is, the amount of variation changes along the slope of a reaction norm. Second, the magnitude of within-individual variation is related to personality, that is, the amount of within-individual variation changes in relation to the average behavioural tendency of an individual. Blum and colleagues could show that these relationships are both non-linear. "Extreme personalities", i.e. individuals that had either very high or very low values in a personality trait, showed less within-individual behavioural variability. Also, situational gradients produced lower among-individual variation at both extreme ends of the reaction norm (see Figure 2). To give an example, we assume that individuals of a species vary in their latency to initiate a flight response when seeing a predator. Let us also assume varying distance of a predator as a situational gradient, low distance would be a dangerous situation and high distance would be less dangerous. Based on the NIPS model, all individuals should show a similar flight response when the distance of the predator is either low or high and vary to a greater extent when distance is at a medium level. That is, when the predator comes into sight but is far away, all individuals will have a high latency, because they can afford to calmly assess the situation. When the predator is near, all individuals will have a short latency, because they all want to save their life quickly. When the predator is at a medium distance, some individuals may play safe and have a short latency, others will take more time to wait whether the predator will clear away. The same pattern could also occur for repeated measurements of the same individual along the reaction norm. This has also been referred to as "situational strength" in the psychology literature.

What the NIPS model does not include are more permanent environmental gradients, which are however often studied by behavioural ecologists. We could include this in the model by saying that flight response does not only vary with how dangerous a specific situation is, but also with how dangerous predator situations are in general in different environments an individual may encounter. Such environmental gradient

could quantify how easily our predator can detect and attack the species we study (high, medium, or low predation environment). This could produce an additional reaction norm, potentially also showing a non-linear relationship with between-individual or within-individual variance. In other words, alongside both reaction norms variance will be heterogeneous, which could violate assumptions of statistical models if not accounted for (Cleasby, Nakagawa, & Schielzeth, 2015, but see Schielzeth et al., 2020). Now a flight response is probably not only based on the situation or a more general environment, but also on characteristics of the individual. The NIPS model posits that individuals who are generally very bold or very fearful show lower variability in their flight response in reaction to a situational gradient than average individuals. We should thus observe milder slopes in the reaction norm of "extreme" individuals and steeper slopes in average individuals. In other words, Blum and colleagues assume a non-linear relationship between personality and the magnitude of plasticity in a behaviour.

As outlined above, we have to be careful when translating this model from one discipline to another. In the NIPS model, "personality" is an additional variable representing values in a latent personality trait (e.g., boldness) which are separate from the values of the behavioural variable (flight response in our example). This means "personality" is included separately in a statistical equation to predict the behaviour. Behavioural ecologists call the between-individual variation in the mean of the behaviour "personality" and will unlikely expect "personality" to be a separate latent variable. A non-linear relationship between personality and the magnitude of plasticity in a behaviour would therefore equate to a non-linear relationship between the mean and the magnitude of plasticity in a behaviour (see Figure 3). In our example, individuals who have a high or low mean flight response would have a lower magnitude of plasticity than the average individual. In behavioural ecology, personality-plasticity relationships have been, for example, discussed as personality-related differences in the sensitivity to environmental variation (Verbeek, Drent, & Wiepkema, 1994; Mathot, Wright, Kempenaers, & Dingemanse, 2012). In psychology, such relationships are also called behaviour-situation contingencies (Fleeson, 2007). Similarly, (non-)linear relationships between measures of an internal state, for example stress responsivity or energy reserves, and plasticity of behaviour may occur (Dingemanse & Wolf, 2013), which could be included in a statistical model as well. Another addition to the

NIPS model would be to consider relationships between the mean and predictability of a behaviour. Mitchell and colleagues (2021) give an example in which bold individuals show lower residual within-individual variance (higher predictability) in flight response behaviour than more fearful individuals because this would be more costly for them (they are at higher risk of being attacked). Contrary, we could think of highly sensitive physiological reactions to predation cues that produce hardly any residual within-individual variation in fearful individuals. Taken together we may thus also encounter non-linear relationships between the mean and magnitude of predictability in behaviour.

The interaction of multiple situational / environmental reaction norms and the mean behaviour of individuals could also lead to hypotheses about when individuals form or deviate from a general pattern of within-individual variability, that is, under which conditions plasticity and predictability are correlated and under which they are not. In our example, we could assume that fearful individuals, that is individuals who have on average a short latency in their flight response, usually show low within-individual variability: they have a low magnitude of plasticity in reaction to fluctuating predation risk in the environment and also low residual within-individual variance in their behaviour when they encounter a predator in most situations (high predictability). Only in the special case of very high distance of the predator in the situational gradient these individuals risk higher residual within-individual variance in their behaviour and deviate from the usual pattern of correlated plasticity and predictability (as Individual 5 in Figure 3 does). This would be an extension of the influence of multi-dimensional reaction norms on predictability as outlined above, as it includes an additional relationship to the individual's mean behaviour. Various other relationships between personality, plasticity and predictability may occur. Here we focussed on a correlation between personality and magnitude of within-individual variance. There may also be, for example, correlations that involve the direction of plasticity (intercept-slope or slope-slope correlations), which we did not cover for reasons of brevity. We refer interested readers to O'Dea (2021). An introduction to the statistical quantification of between-individual differences in within-individual variance can be found in Cleasby and colleagues (2015). In extension to previous reviews (e.g., Carter et al., 2013), we give an updated glossary of how terms may be translated between disciplines in Table 1.

Page 56

4. Perspectives on correlated traits

Apart from research focus and terminology, a difference between disciplines is how correlated behaviours are usually perceived and analysed. We pointed out repeatedly that psychologists typically study individual differences in global traits (which are also called dimensions or factors). We also outlined that researchers historically aggregated correlated behaviours to more global traits in order to simplify test administration or subsequent statistical analyses. Although this could be considered a mere practical necessity, there is also a longstanding belief in psychology that global traits reflect a latent common cause or general disposition within individuals that affects the expression of various behaviours and therefore produces correlations among them. This explanation is not without alternatives, however. Here we integrate perspectives on the occurrence of suites of correlated behaviours by both disciplines.

4.1 A behavioural ecology perspective

The observation of between-individual correlations of behaviours has been introduced to the behavioural ecology literature as *behavioural syndromes* (Sih, Bell, & Johnson, 2004). A between-individual correlation refers to the correlation of individuals' mean behaviour for two or more traits. This is important to note, because if we calculate the correlation between behaviours measured once, the resulting coefficient could also reflect covariation within individuals. A within-individual correlation exists when an individual's change in one behaviour between time period *t* and *t + 1* is correlated with its change in another behaviour over the same period (Dingemanse & Dochtermann, 2013). This could occur, for example, because individuals show a similar plastic response to variation in the environment in two behaviours (joint plasticity). To measure a behavioural syndrome, it is therefore necessary to decompose the raw *phenotypic* correlation into between- and within-individual components, which requires multiple measurement occasions (Dingemanse, Dochtermann, & Nakagawa, 2012). If correlations within individuals differ from the correlations between individuals, but every behaviour is only measured once, the resulting un-partitioned phenotypic correlation could be a gross misestimate of the between-individual correlation (Niemelä &

Dingemanse, 2018), although this was not necessarily the case in empirical studies (Brommer & Class, 2017). The separation of between- and within-individual correlations is important, because both can be caused by environmental effects, but only between-individual correlations can reflect additive genetic correlations (Brommer, 2013). Behavioural syndromes that reflect a genetic correlation between traits are in opposition to earlier work arguing against the occurrence of correlated behaviours (Wilson 1998; Coleman & Wilson, 1998). They constrain the possibility of calibrating traits independently to environmental challenges (Dingemanse & Dochtermann, 2013). Why, then, do they occur? The reason could be, that behavioural syndromes are adaptive themselves, that is, the result of selection in a specific environment. If so, they should occur or vanish depending on the environment a population occupies or because of niche-specialization within populations (Bergmüller & Taborsky, 2007). An example is how bold, active, and aggressive behaviour is more tightly (i.e., higher) correlated in populations of sticklebacks with high predation risk than in populations with low predation risk (Bell, 2005; Dingemanse et al., 2007; Brydges et al., 2008). Notably, these were also correlations of behaviours shown in different contexts, that is, correlations between aggressive behaviour towards conspecifics, explorative behaviour in an unfamiliar environment and foraging behaviour under risk. An explanation for such cross-context correlations would be that more solitary fish, monopolizing their own patch of a habitat, may need to be more explorative due to the relative lack of social information to find such patch and at the same time more aggressive towards conspecifics to defend their patch, once they found one (Dingemanse et al., 2007). Possibly, the tighter phenotypic correlations in the high-risk environment could also be tighter on the genetic level. This was however not the case, since the genetic correlation matrix of behavioural traits was largely stable across populations (Dingemanse, Barber, & Dochtermann, 2020). This means that the difference in phenotypic correlations among environments is likely a product of a joint plastic adjustment of multiple behaviours ("plasticity integration") due to unique aspects of the high-risk environment, which were however not measured in the experiment.

Findings of cross-context and temporally stable behavioural syndromes (which are reported for various species, see e.g., Boulton et al., 2014; Garamszegi, Markó, & Herczeg, 2012; Perals Griffin, Bartomeus, &

Sol, 2017) reflect a correlation matrix similar to correlated behaviours that are represented by personality dimensions in psychology. Also, various studies in behavioural ecology include latent variables (e.g., Boulton et al., 2014; Dochterman & Jenkins, 2007, Royauté, Buddle, & Vincent, 2014). When do behavioural ecologists consider it reasonable to model behavioural syndromes as latent variables, like psychologists do? Since research in behavioural ecology is rooted in trying to understand evolutionary aspects of behavioural (co-)variation, this question leads to a concept in evolutionary biology termed phenotypic integration (Pigliucci, 2003). Phenotypic integration occurs when evolutionary selection favours a correlation among different traits. A classic example is how different parts of the cranium are correlated in size ("integrated") to form a functioning system (Cheverud, 1996). Deviations from this integration of sizes will lead to a loss of functionality and fitness, which is why a correlation among them will be selected for (Cheverud, 1982; Schwenk & Wagner, 2001). Such fitness interaction among traits can produce genetic correlations among them, for example by creating linkage disequilibrium (i.e. non-random association of alleles at different genetic loci, Brodie, 1992; Cheverud, 1982). We can extend this example to behavioural traits, when a specific combination of trait values works better (i.e., produces higher fitness, is more adaptive) than other combinations of trait values in a certain environment, leading to correlated selection (Brodie, 1992; Endler, 1995; Sinervo & Svensson, 2002).

Note that although many systems of correlated traits reflect genetic correlations (Roff & Fairbain, 2012), potentially because of their joint effect on fitness, genetically unrelated traits may also covary, for example because they are both, but independently, affected by the same environmental factor (Endler, 1995). Also, genetic correlations could be due to pleiotropy (a single gene affecting multiple behaviours), genetic linkage (genes closely together on the same chromosome) or shared physiological mechanisms. These are stronger constraints on independent behavioural adaptation than selection-induced genetic correlations due to linkage equilibrium, because the latter can break down more easily and at a faster rate (when the association of alleles is no longer selected for; Royauté, Hedrick, & Dochtermann, 2020). Alike the correlated size of parts of the cranium, correlated behaviours that interact in their effect on fitness may be seen as a functional system or unit. This sense of unity compares to psychologists' view of functionally

related behaviours belonging to a joint personality dimension, even though the nature of these latent personality variables is usually less explicitly defined. The concept of phenotypic integration provides behavioural ecologists a theoretical foundation based on which the use of latent variables can be embedded in evolutionary theory. For example, aggressiveness in great tits can be seen as an evolved unit represented by a latent variable, serving the function of displacing intruders (Araya-Ajoy & Dingemanse, 2014). Likewise the genetic covariance among predation-related behaviours in the example of sticklebacks above was modelled in a structural equation model as latent variable (Dingemanse et al., 2020).

4.2 A personality psychology perspective

When psychologists started to study individual differences in behaviour, correlations among them were not questioned, since they were obviously present in the data of self-descriptions and there was also no theory opposing to this. Instead they were welcome for practical purposes, as they allowed for forming broader personality dimensions that predicted life outcomes such as educational and occupational success. The question of why they were present primarily focussed on finding proximate explanations of underlying mechanisms (see, e.g., Eysenck & Levey, 1972; Depue & Collins, 1999) and then fanned out to cover also evolutionary questions (Penke, Denissen, & Miller, 2007). While going the path from a somewhat different direction, psychologists have addressed similar questions about behavioural variation as behavioural ecologists did. Today we know that phenotypic correlations among questionnaire items of the very common Five Factor Model (FFM) of human personality are mirrored by genetic correlations and also fairly well by the residual environmental correlations. Factor analyses of the respective correlation matrices produce very similar latent variable structures for phenotypic, genetic and environmental correlations across populations (Yamagata et al., 2006).

Concerning the genetic correlations among FFM items that load on a common factor we consider it important to note, that they do not so because of a specific genetic mechanism. Rather, as posited by the phenotypic null hypothesis of behaviour genetics, every trait is heritable and the multivariate factor struc-

ture we find on the phenotypic and genetic level is due to a general and unsystematic genetic background of correlated behaviours (Turkheimer, Pettersson, & Horn, 2014). Concerning environmental influences, we also note that the similarity of latent factor structures applies mostly to western, educated, industrialized, rich, and democratic populations (described by the acronym WEIRD; Henrich, Heine, & Norenzayan, 2010), which most psychological studies are based upon. Although phenotypic covariance patterns show overall considerable cross-cultural consistency (Church & Lonner, 1998; Rolland, 2002; Thalmayer & Saucier, 2014), they do not equally apply to all human populations. We know, for example, that personality models such as the FFM do not replicate in some indigenous or rural, largely preliterate, populations (Gurven et al., 2013, Rossier et al., 2013), which are under-represented in cross-cultural studies and reviews (Church, 2016).

Empirical and simulation studies (Lukaszewski et al., 2017; Smaldino et al., 2019; Durkee et al., 2022) link varying behavioural correlations between human populations to differences in niche diversity. This integrates with variation of behavioural covariation patterns between non-human animal populations (e.g., Bell, 2005; Dingemanse et al., 2007; Moretz, Martins, & Robison, 2007; Michelangeli et al., 2018) and the literature on behavioural variation due to (social) niche specialisation in behavioural ecology (Bergmüller & Taborsky, 2007, Bergmüller & Taborsky, 2010; Laskowski & Bell, 2014, Montiglio, Ferrari, Réale, 2013). Findings from humans show that higher socioecological complexity and niche diversity is associated with lower covariation amongst personality factors and behavioural traits in general, thereby producing also a larger number of factors in a factor analysis (Lukaszewski et al., 2017; Smaldino et al., 2019; Durkee et al., 2022). The theoretical rationale behind these findings is that greater niche diversity allows for greater variance of behavioural traits and covariance patterns amongst them (we need to be careful with causality here though, since active gene-environment correlations could also increase niche diversity through niche construction; Saltz, 2019).

Personality research has for a long time conceptualized latent personality factors as a common cause to variation in associated behaviour (Borsboom et al., 2003). It seems however farfetched to assume that human populations in which the FFM does not hold true have a unique set of common causes to behavioural covariation. Rather, the correlational structure among FFM items is probably due to a mixture of different causes. For example, population differences could be the result of variation in fitness outcomes of behavioural trait combinations in different (social) environments. Higher-order latent variables could thus be partly constituted of general strategies that involve coordinated expression of multiple traits. If this were true, personality factors could be maintained by negative frequency-dependent selection on such broader strategies. This would lead to correlated means (intercepts) in the behavioural reaction norm (what behavioural ecologists call a behavioural syndrome). If a coordinated response to environmental variation among behaviours is favoured, this would also lead to correlated slopes (plasticity syndrome). The genetic correlations underlying personality factors may thus partly occur due to correlated selection on heritable behaviours (as also described in behavioural ecology, see above) and not necessarily due to a specific genetic mechanism that affects multiple traits.

On the other hand, an understanding of latent personality factors as general dispositions that guide a manifold of behaviour must not be dismissed entirely. An explanatory entry to such view is provided by endophenotypes that affect behavioural covariation among and within individuals. Endophenoytpes are heritable intermediate phenoytpes in the causal chain between genetic and behavioural variation, for example dopaminergic reward responses or amygdala reactivity (Penke, Denissen, Miller, 2007; Wacker & Smilie, 2015). As for behaviour or more general behavioural strategies, selection could maintain individual variation in the reactivity of endophenotypes and associated latent personality factors. We can also think of fitness benefits and costs of variation in such endophenotypes (see Nettle, 2006; Nettle & Penke, 2010) and also that such benefits and costs vary between populations. An example would be that in less individualized societies (e.g., where people live in extended family clusters), personality domains as extraversion and agreeableness mix into a general prosocial disposition because the benefits of being extraverted are conditional on agreeable manners (e.g., people are valued who talk confidently but modestly, Gurven et al., 2013). Hy-

pothetically, this could lead to a tighter correlation of endophenotypes that produces a lower number of in-dependent latent personality factors (Gurven et al. 2013; Penke et al., 2007) and population differences among behavioural correlation matrices.

Research on the item-level of personality questionnaires strengthens our belief in a mixture of contributing causes to latent personality factors. Consider that behaviours or questionnaire items are not perfectly correlated to latent factors, they usually harbour residual variation (i.e. variation not explained by the factor). The idiosyncratic residual variation of single behaviours or items is known to be heritable and a valid predictor of various associated outcomes as well (Mõttus, 2016; Mõttus, Kandler, Bleidorn, Riemann, McCrae, 2017). Also, causality of a latent factor on the associated items would imply that a change in the state of the factor, i.e. an individual adapts to environmental change on the level of the factor, leads to concerted change in associated items in a way that within-individual correlations of repeated measurements mirror among-individual correlations (Borsboom et al., 2003; Molenaar & Campbell, 2009). This is an old argument: Cattell noted in 1946 that traits which represent functional unity should show common variation amongst their indicators in fluctuations within individuals. In other words, behavioural correlations should be ergodic. Aggressiveness in great tits could be an example of functional unity, as correlations among behaviours were in fact shown to be ergodic (Araya-Ajoy & Dingemanse, 2014). From what we know about the factors of the FFM, they are not, meaning that they differ on the among- and within-individual level (Molenaar, 2004).

In summary, between-individual correlations of behaviours / questionnaire items could be partly due to a common cause (e.g., variation in underlying endophenotypes) and partly due to specific environmental influences or dependencies that result in a correlated trait expression. If we further perceive endophenotypes as functional systems that are genetically integrated (Schwenk & Wagner, 2001) and evolutionary conserved (e.g., the HPA-axis, Denver, 2009), they would contribute to the overall observed stability in the genetic covariance matrix across populations and constrain independent variation of associated behaviour. Pheno-

typic correlations among behaviours could then be the result of such proximal constraints plus adaptive joint expression of behaviour sensu correlated selection (see above), which in turn could be shared between populations (potentially adding to shared genetic covariance across populations) or be population-specific. This would be a mixture of what has been termed the "adaptive hypothesis" and the "constraint hypothesis" of behavioural syndromes in behavioural ecology (Royauté et al., 2020) and relate to the mixture of pleiotropic effects and linkage disequilibrium in the genetic architecture of human personality (Polushina et al., 2021; Vinkhuyzen, 2012).

4.3 The Common-Cause-Reliability-Jumble

In Section 2 we briefly pointed out that residual within-individual variance is sometimes referred to as random measurement error in psychology, though "error" is merely a name given to the ignorance of the many forces that influences an individuals response in a specific event (Cronbach, 1947). Many psychological studies try to quantify how "error-free" a measurement is, which is called measurement *reliability* in psychological test theory. Various metrics of reliability have been developed, for example by correlating parallel versions of a test measuring the same attribute or by computing the *repeatability* of individual test scores, that is, the extent to which repeated measurements of individuals produce the same test results (Sijtsma, 2009). If the distribution of repeated test scores shows zero variation within individuals, the test is perfectly repeatable (i.e., reliable). Although behavioural ecologists also calculate repeatability as a standard routine, they do not use this statistic as a proxy of measurement reliability, because residual variance due to influences that are not included in the model is not considered "error" (e.g., unmeasured variation in plasticity to hidden variables). Experimental animal studies reveal how difficult it is to control variation in behavioural tests. For example, Bierbach, Laskowski, and Wolf (2017) report a repeatability of 0.35 in a standardized behavioural test measuring activity patterns of genetically identical clonal fish raised in nearly identical rearing conditions. This suggests a fairly high proportion of within-individual variation of activity in a highly standardized set-up where investigators attempted to reduce environmental variation entirely. Similar res-

ults have been reported for genetically identical fruit flies raised under identical conditions (Kain, Stokes, & de Bivort, 2012).

Because of the difficulty of administering independent repeats of the same test, where variation within individuals is only due to random error, psychological studies often calculate reliability based on multiple indicators of an attribute measured at one point in time. The rationale behind this approach is that every indicator, for example every questionnaire item, is considered an independent assessment of the same attribute (Guttman, 1945). Following this rationale, random measurement errors in single indicators are thought to balance each other out to an asymptotical value of zero when the single indicators are aggregated, which will lead to an increase in *concurrent reliability* (Epstein, 1983). Most measures of concurrent reliability reflect the proportion of variance in the multiple indicators that is due to an error-free common cause, conceptualized and in some cases also computed based on a latent factor underlying these indicators (see e.g. Sijtsma, 2009; Geldhof et al., 2014; Trizano-Hermosilla & Alvarado 2016; McNeish, 2017). The logic here is that if shared variance among indicators is high, measurement error must be low.

As we argued above, the common cause assumption is a fairly strong one and probably not met in many associations of complex behaviours. Latent variables yield higher reliability estimates than single behavioural measurements (Furr & Funder, 2004, Rushton, Brainerd, & Pressley, 1983), but next to measurement error we may also lose meaningful variation unique to the behaviours we summarize into a latent variable score. As latent variables are hardly used in behavioural ecology, measures of concurrent reliability (e.g., Cronbach's alpha, McDonald's Omega) are typically not used either. Since (pure) latent variable models are also challenged in psychology, it seems timely to refrain from using such estimates as measures of "reliability" entirely. Nevertheless they remain a possibility of estimating consistency, if such is of interest. An applicable alterbative to both behavioural ecology and psychology is to include information that potentially leads to "noise" as additional variables in thestatistical model (e.g., different measuring devices, different times of the day an individual was tested, etc., Nakagawa & Schielzeth, 2010). This produces what behavi-

oural ecologists call estimates of *adjusted repeatability,* where variation between and within individuals is tried to be reduced to meaningful "biological" variation.

5. General discussion

With our review we aimed to further connect research on "personality" as a whole, that is, to tear down boundaries that emerged due to a separation of disciplines and their scientific journals. We observed an emerging interdisciplinary consensus on various topics in this research area, which we will summarize and discuss here jointly.

We outlined above that behavioural ecologists and personality psychologists uncovered different violations of assumptions underlying a "pure" latent trait perspective on correlated behavioural differences. The belief that such correlations occur solely due to a common cause is hence vanishing. Among these violations are that "personality factors" are often non-ergodic (correlations between behaviours differ on the between- and within-individual level), which means that individuals do not change their behaviour in a concerted way as predicted by between-individual correlations. A single common cause is also unlikely considering examples where environmental variation between different populations produces different phenotypic correlations between behaviours, whereas genetic correlations are stable, or that residual variances of behaviours which are not accounted for by a latent trait covary and predict life outcomes in meaningful ways as well.

Following the disputable validity of latent personality traits, the common practice in personality psychology to assess measurement reliability by computing cross-sectional, concurrent indices based on the correlation matrix among behaviours / questionnaire items (e.g., Cronbach's alpha, McDonald's Omega) needs to be re-thought as well. Behavioural ecologists never attempted to do so.

Within-individual variation in behavioural expression is of very similar magnitude in humans and non-human animals. An average repeatability of behaviour around approximately 0.40 indicates that the larger part of total behavioural variation happens within individuals (as compared to between individuals). On the other hand, behavioural ecologists failed to suppress between-individual variation in studies of genetically cloned individuals under highly standardized rearing conditions, possibly indicating that random (i.e., non-trackable) developmental processes are an important source of between-individual variation as well.

Findings from behavioural ecology and personality psychology consistently support a link between environmental diversity and behavioural variation within and across populations. It has been shown that a diversity of niches within an environment can explain the maintenance of inter-individual differences in behaviour of members of the same population and that populations differ in how tightly various behaviours are integrated based on differences in niche diversity allowing for diversity in behaviour.

The behavioural response to environmental challenges will likely be constrained by the inherited genetic architecture of an individual and the interplay between genetic architecture and environment. We can certainly think of scenarios where individuals have not much choice or influence over environmental conditions and challenges (e.g., fluctuating resource abundance, Dingemanse et al., 2004), but for the most cases we have to expect that they actively seek out or constructs their own niche ("niche picking" and "niche construction"; see e.g., Stamps & Groothuis 2010; Buss & Plomin, 1986; Penke, 2010). In other words, we assume that individuals carry a certain genetic disposition that affects the average tendency in their behaviour for which they seek a good fit in the environment. This is called active gene-environment correlation (rGE), which is a common explanation of personality variation in psychology and human behaviour genetics (Johnson & Penke, 2014). Also, we cannot expect an independent response of a single behaviour if we assume that multiple behaviours are correlated on the genetic level.

Many studies in behavioural ecology and personality psychology do not simultaneously study the multitude of relationships among various components in behaviours. For every single behaviour we could be interested in the variation of average tendencies, plasticity, or predictability between individuals. For every single behaviour the average value may be correlated with plasticity or predictability in this behaviour and also plasticity and predictability may be correlated within a single behaviour. Covariation may also occur between two or more behaviours, be it covariation between average values, between plasticity of both behaviours, or their predictability. Finally, covariation may not only occur between, but also within individuals, which we have to account for by taking multiple measurements.

As we have exemplified for the NIPS model above, relationships between variance components of a behaviour, between different behaviours, and between behaviour and environment need not to be linear. Rather, we would expect, for example, a non-linear relationship between an individuals average value and plasticity of a behaviour, where individuals with very low or very high average behavioural tendencies are less plastic.

Relationships between variance components of behaviour or between behaviour and environment may change throughout ontogeny or over evolutionary timescales. An example would be that individuals are more limited in their ability or possibility to seek out niches at earlier ages than later on in life. For humans, there is some evidence that personality differences among individuals increase from infancy to adulthood, partly due to amplified additive genetic differences, indicating gene-environment correlations, but also due to an increase of non-additive genetic effects (Mõttus et al., 2019). A possible explanation for the latter finding is that behaviours interactively influence each other during ontogeny, as posited by advocates of process-orientated and network models of personality.

Network models account for instantiations where correlations between behaviours are formed through logically linked dependencies or interactions between them (e.g., liking people and going to parties, Cramer et al., 2012), rather than by an underlying global trait that acts as a unidirectional common cause. The estimated network is a graphical representation of the syndrome structure among variables which may arise solely due to feedback loops of sequentially linked behaviours (e.g., sleeping problems lead to concentration problems, which lead to trouble in school, which leads to worrying, which leads to sleeping problems). Network models are graphical representations of what many behavioural ecologists report in their studies, that is, covariation among behaviours and potential covariates among and within individuals which are estimated by multivariate linear mixed models. Similar to network models in psychology, behavioural ecologist Endler (1995) pointed out how trait correlations in animal species could occur when the function of one trait is required before the other is performed. That is not to say this is the only possible pathway. Endler also gave the example that genetic variation in pituitary function affects multiple traits in stickleback. In generalized network models (GNM) we may combine the dynamics of such common causes of behaviours (latent variables) and behavioural dynamics independent of a common cause (networks of residuals that cannot be explained by latent variables; Epskamp, Rhemtulla, & Borsboom, 2017; Guyon et al., 2017). This integration extends to any covariates of interest. In medicine, integrative network models of, for example, behavioural symptoms, polygenic scores and metabolic markers, are termed multi-plane or multi-layer networks (Guloksuz, Pries, & van Oz, 2017). Process models of personality development integrate with such network perspective (Geukes, van Zalk, & Back, 2018). Here, between-individual differences in genetic background are argued to provide a set-point of average expression across behaviours, but the correlated expression of behaviours in a specific situation ("state expression") is further understood as part of a process of related states in goals, strategies, experiences, and evaluations thereof. That is, such states result from sequentially linked dependencies, rather than a common cause. Consider how a specific goal may lead to a strategy to reach that goal, which will in turn affect the correlated expression of various behaviours, subsequent experiences, and mental evaluations of these experiences. If an individual changes in the expression of these behaviours, this may be due to a change in any of the domains along the process chain: The individual could have changed the goal, or the strategy, could have made unexpected experiences, or

could have changed in the evaluation of these experiences, which may lead to differential expression of behaviour in the future (Geukes et al., 2018). If a change in any of the process domains stabilizes, the mean expression of related behaviours may move away from the genetically influenced set-point. We can transfer such process model to behavioural ecology and the behaviour of non-human animals, where, for example, (co-)variation in behavioural expression is linked to variation in foraging strategies, fluctuating success of these strategies, and adaptation of foraging strategies. As behavioural ecologists never really adopted the perspective of a single causal pathway and analyse their data predominantly with (generalized) linear mixed models (Dingemanse & Wright, 2020), network or process models do not offer a paradigm shift, but rather an extension of how results can be visualized or how latent variables can be additionally included if warranted.

Since behavioural ecologists and personality psychologists agree that an estimate of a behavioural disposition or tendency of individuals can be derived as the average of repeatedly measured behavioural instantiations, we are in line with Roberts (2018) to propose that future studies in personality psychology will benefit from measuring only "personality states" instead of retrospective mental aggregations that potentially conflate what has been termed "personality state" and "personality trait" in psychology, whenever possible.

Table 1: Glossary of terms across disciplines

| Term / Concept	Psychology	Behavioural Ecology
Personality	An individual's unique mixture of varying degrees in traits that encompass behaviour, feelings, and thoughts	Among-individual variation in the mean of a behaviour in the average environment
Trait	A temporally stable characteristic of an individual	A characteristic of an individual
Personality trait	A (usually) global trait that describes variation in correlated behaviours, feelings, and thoughts	A behaviour in which individuals show consistent among-individual differences = essentially every behaviour
Personality state	Short-term deviation from the average personality trait value of an individual	Term not used, could be any behavioural measurement
Repeatability	Term seldom used, equates to consistency of among-individual differences measured via an intra-class coefficient (ICC)	Amount of variance due to among-individual differences in relation to total phenotypic variance
Reliability	Refers to how „error-free" / reproducible measurements are, possibly quantified by various different coefficients	Term not used in addition to repeatability
Plasticity	Within-individual variation in personality factor scores or behaviour, term seldom used in this area of research	Within-individual variation in a given trait in relation to variation in the environment
Predictability	Term seldom used, could be used as in behavioural ecology	Refers to residual within-individual variation in a trait that is not accounted for by a statistical model (could be due to measurement error or hidden influence of additional variables)
Behavioural Syndrome	Term seldom used, in clinical psychology a syndrome refers to correlated symptoms of a mental disorder	Correlation of among-individual variance between two or more behaviours
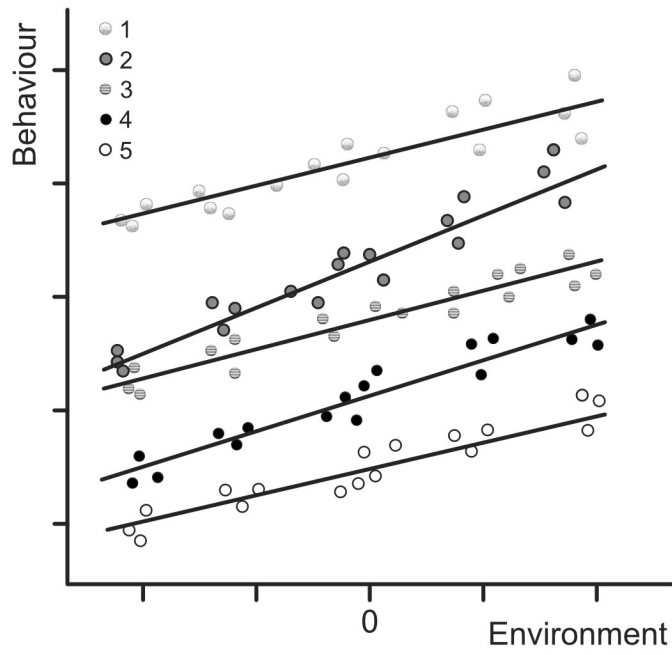
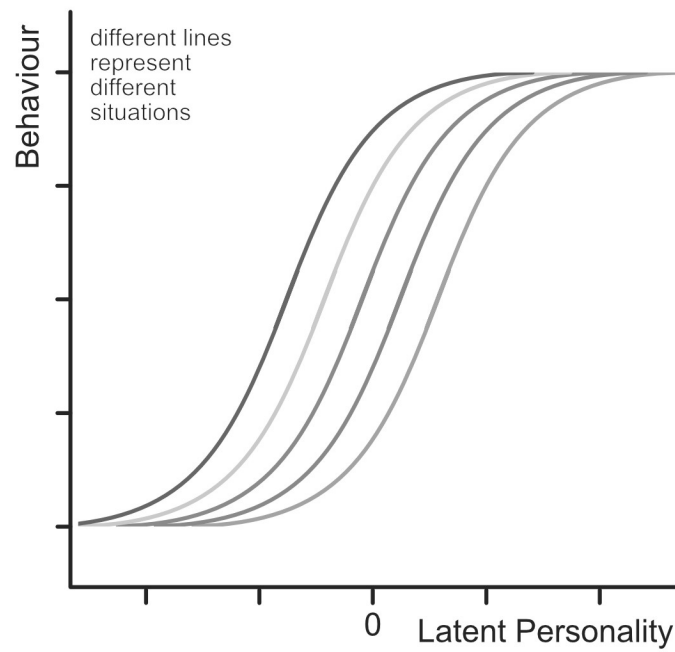Figure 1: Behavioural reaction norms for five individuals

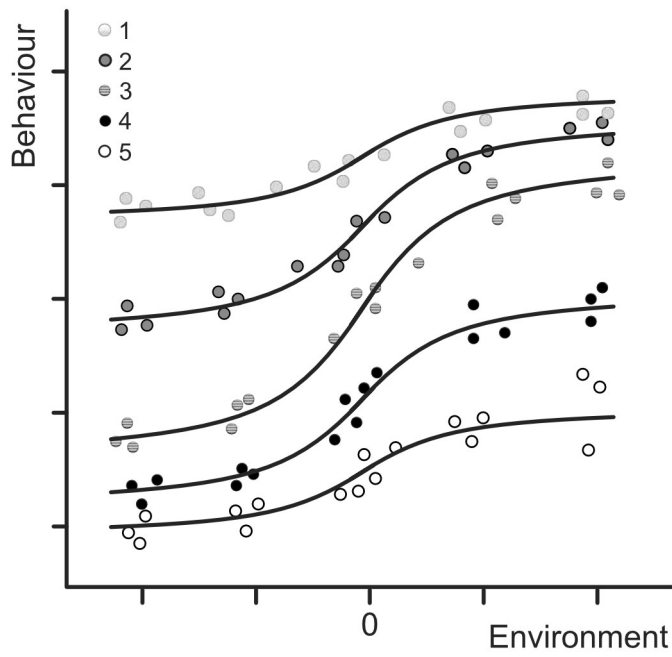Figure 2: Behavioural reaction norm for five situations as posited by the NIPS model

Figure 3: Behavioural reaction norm for five individuals adapted from the NIPS model, but without a latent personality variable

References

Adams, M. J., Majolo, B., Ostner, J., Schülke, O., De Marco, A., Thierry, B., Engelhardt, A., Widdig, A., Gerald, M. S., & Weiss, A. (2015). Personality structure and social style in macaques. *Journal of Personality and Social Psychology*, *109*(2), 338.

Allport, F. H., & Allport, G. W. (1921). Personality Traits: Their Classification and Measurement. *The Journal of Abnormal Psychology and Social Psychology*, *16*(1), 6.

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological monographs*, *47*(1), i.

Araya-Ajoy, Y. G., & Dingemanse, N. J. (2014). Characterizing behavioural „characters": An evolutionary framework. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1776). https://doi.org/10.1098/rspb.2013.2645

Araya☐Ajoy, Y. G., & Dingemanse, N. J. (2017). Repeatability, heritability, and age☐dependence of seasonal plasticity in aggressiveness in a wild passerine bird. *Journal of Animal Ecology*, *86*(2), 227–238.

Back, M. D., Schmukle, S. C., & Egloff, B. (2009). Predicting Actual Behavior From the Explicit and Implicit Self-Concept of Personality. *Journal of Personality and Social Psychology*, *97*(3), 533–548. https://doi.org/10.1037/a0016229

Bateson, P., & Laland, K. N. (2013). Tinbergen's four questions: An appreciation and an update. *Trends in ecology & evolution*, *28*(12), 712–718.

Baumgarten, F. (1933). *Die Charaktereigenschaften*. A. Francke.

Beck, A. T., & Haigh, E. A. P. (2014). Advances in cognitive theory and therapy: The generic cognitive model. *Annual review of clinical psychology*, *10*, 1–24.

Bell, A. M. (2005). Behavioural differences between individuals and two populations of stickleback (Gasterosteus aculeatus). *Journal of Evolutionary Biology*, *18*(2), 464–473. https://doi.org/10.1111/j.1420-9101.2004.00817.x

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

Bergmüller, R., & Taborsky, M. (2007). Adaptive behavioural syndromes due to strategic niche specialization. *BMC ecology*, 7(1), 1–7.

Bergmüller, R., & Taborsky, M. (2010). Animal personality due to social niche specialisation. *Trends in ecology & evolution*, 25(9), 504–511.

Bierbach, D., Laskowski, K. L., & Wolf, M. (2017). Behavioural individuality in clonal fish arises despite near-identical rearing conditions. *Nature communications*, 8(1), 1–7.

Bleidorn, W. (2009). Linking personality states, current social roles and major life goals. *European Journal of Personality*, 23(6), 509–530.

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological bulletin*, 117(2), 187.

Blum, G. S., Rauthmann, J. F., Göllner, R., Lischetzke, T., & Schmitt, M. (2018). The Nonlinear Interaction of Person and Situation (NIPS) Model: Theory and Empirical Evidence. *European Journal of Personality*, 32(3), 286–305. https://doi.org/10.1002/per.2138

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

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2003). The Theoretical Status of Latent Variables. *Psychological Review*, 110(2), 203–219. https://doi.org/10.1037/0033-295X.110.2.203

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(5), 791–803.

Brodie III, E. D. (1992). Correlational selection for color pattern and antipredator behavior in the garter snake Thamnophis ordinoides. *Evolution*, *46*(5), 1284–1298.

Brommer, J. E. (2013). Variation in plasticity of personality traits implies that the ranking of personality measures changes between environmental contexts: Calculating the cross-environmental correlation. *Behavioral Ecology and Sociobiology*, *67*(10), 1709–1718.

Brommer, J. E., & Class, B. (2017). Phenotypic correlations capture between-individual correlations underlying behavioral syndromes. *Behavioral ecology and sociobiology*, *71*(3), 50.

Brydges, N. M., Colegrave, N., Heathcote, R. J. P., & Braithwaite, V. A. (2008). Habitat stability and predation pressure affect temperament behaviours in populations of three-spined sticklebacks. *Journal of Animal Ecology*, *77*, 229–235. https://doi.org/10.1111/j.1365-2656.2008.01343.x

Buirski, P., Kellerman, H., Plutchik, R., Weininger, R., & Buirski, N. (1973). A field study of emotions, dominance, and social behavior in a group of baboons (Papio anubis). *Primates*, *14*(1), 67–78.

Buss, A. H., & Plomin, R. (1986). The EAS approach to temperament. *The study of temperament: Changes, continuities and challenges*, 67–79.

Carter, A. J., Feeney, W. E., Marshall, H. H., Cowlishaw, G., & Heinsohn, R. (2013). Animal personality: What are behavioural ecologists measuring? *Biological Reviews*, *88*(2), 465–475.

Caspi, A., Roberts, B. W., & Shiner, R. L. (2005). Personality development: Stability and change. *Annual Review of Psychology*, *56*, 453–484. https://doi.org/10.1146/annurev.psych.55.090902.141913

Cattell, R. B. (1943). The description of personality: Basic traits resolved into clusters. *The journal of abnormal and social psychology*, *38*(4), 476.

Cattell, R. B. (1946). *Description and measurement of personality.*

Chamove, A. S., Eysenck, H. J., & Harlow, H. F. (1972). Personality in monkeys: Factor analyses of rhesus social behaviour. *Quarterly Journal of Experimental Psychology*, *24*(4), 496–504.

Cheverud, J. M. (1982). Phenotypic, genetic, and environmental morphological integration in the cranium. *Evolution*, 499–516.

Cheverud, J. M. (1996). Developmental integration and the evolution of pleiotropy. *American Zoologist*, *36*(1), 44–50.

Church, A. T. (2016). Personality traits across cultures. *Current Opinion in Psychology*, *8*, 22–30.

Church, A. T., & Lonner, W. J. (1998). The cross-cultural perspective in the study of personality: Rationale and current research. *Journal of cross-cultural psychology*, *29*(1), 32–62.

Clark, A. B., & Ehlinger, T. J. (1987). Pattern and adaptation in individual behavioral differences. *Perspectives in ethology*, 1–47.

Cleasby, I. R., Nakagawa, S., & Schielzeth, H. (2015). Quantifying the predictability of behaviour: Statistical approaches for the study of between☐individual variation in the within☐individual variance. *Methods in Ecology and Evolution*, *6*(1), 27–37.

Coleman, K., & Wilson, D. S. (1998). Shyness and boldness in pumpkinseed sunfish: Individual differences are context-specific. *Animal behaviour*, *56*(4), 927–936.

Costantini, G., Epskamp, S., Borsboom, D., Perugini, M., Mõttus, R., Waldorp, L. J., & Cramer, A. O. J. (2015). State of the aRt personality research: A tutorial on network analysis of personality data in R. *Journal of Research in Personality*, *54*, 13–29.

Costantini, G., Richetin, J., Preti, E., Casini, E., Epskamp, S., & Perugini, M. (2019). Stability and variability of personality networks. A tutorial on recent developments in network psychometrics. *Personality and Individual Differences*, *136*, 68–78.

Cramer, A. O. J., van der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., Kendler, K. S., & Borsboom, D. (2012). Dimensions of normal personality as networks in search of equilibrium: You can't like parties if you don't like people. *European Journal of Personality*, *26*(4), 414–431. https://doi.org/10.1002/per.1866

Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: The seven pillars of RDoC. *BMC Medicine*, *11*(1), 126. https://doi.org/10.1186/1741-7015-11-126

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(8), 734–739.

Denver, R. J. (2009). *Structural and functional evolution of vertebrate neuroendocrine stress systems*.

Depue, R. A., & Collins, P. F. (1999). Neurobiology of the structure of personality: Dopamine, facilitation of incentive motivation, and extraversion. *Behavioral and Brain Sciences*, *22*(3), 491–517. https://doi.org/10.1017/S0140525X99002046

Digman, J. M. (1990). *PERSONALITY STRUCTURE: EMERGENCE OF THE FIVE-FACTOR MODEL* (S. 417–457).

Dingemanse, N. J., Barber, I., & Dochtermann, N. A. (2020). Non□consumptive effects of predation: Does perceived risk strengthen the genetic integration of behaviour and morphology in stickleback? *Ecology letters*, *23*(1), 107–118.

Dingemanse, N. J., Both, C., Drent, P. J., & Tinbergen, J. M. (2004). Fitness consequences of avian personalities in a fluctuating environment. *Proceedings of the Royal Society B: Biological Sciences*, *271*(1541), 847–852. https://doi.org/10.1098/rspb.2004.2680

Dingemanse, N. J., & Dochtermann, N. A. (2013). Quantifying individual variation in behaviour: Mixed□ effect modelling approaches. *Journal of Animal Ecology*, *82*(1), 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*(11), 1543–1548.

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*(2), 81–89.

Dingemanse, N. J., & Wolf, M. (2010). Recent models for adaptive personality differences: A review. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1560), 3947–3958. https://doi.org/10.1098/rstb.2010.0221

Dingemanse, N. J., & Wolf, M. (2013). Between-individual differences in behavioural plasticity within populations: Causes and consequences. *Animal Behaviour*, *85*(5), 1031–1039.

Dingemanse, N. J., & Wright, J. (2020). Criteria for acceptable studies of animal personality and behavioural syndromes. *Ethology*, *126*(9), 865–869.

Dingemanse, N. J., Wright, J., Kazem, A. J. N., Thomas, D. K., Hickling, R., & Dawnay, N. (2007). Behavioural syndromes differ predictably between 12 populations of three□spined stickleback. *Journal of Animal Ecology*, *76*(6), 1128–1138.

Dochtermann, N. A., & Jenkins, S. H. (2007). Behavioural syndromes in Merriam's kangaroo rats (Dipodomys merriami): A test of competing hypotheses. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1623), 2343–2349.

Durkee, P. K., Lukaszewski, A. W., von Rueden, C. R., Gurven, M. D., Buss, D. M., & Tucker-Drob, E. M. (2022). Niche Diversity Predicts Personality Structure Across 115 Nations. *Psychological Science*, *33*(2), 285–298. https://doi.org/10.1177/09567976211031571

Edition, F. (2013). Diagnostic and statistical manual of mental disorders. *Am Psychiatric Assoc*, *21*.

Ellis, A. (1991). The revised ABC's of rational-emotive therapy (RET). *Journal of Rational-Emotive and Cognitive-Behavior Therapy*, *9*(3), 139–172.

Endler, J. A. (1995). Multiple-trait coevolution and environmental gradients in guppies. *Trends in ecology & evolution*, *10*(1), 22–29.

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, *50*(1), 195–212. https://doi.org/10.3758/s13428-017-0862-1

Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika*, *82*(4), 904–927.

Epstein, S. (1983). Aggregation and beyond: Some basic issues on the prediction of behavior. *Journal of Personality*, *51*(3), 360–392.

Eysenck, H., & Levey, A. (1972). Conditioning, introversion-extraversion and the strength of the nervous system. *Biological bases of individual behavior*, 206–220.

Feaver, J., Mendl, M., & Bateson, P. (1986). A method for rating the individual distinctiveness of domestic cats. *Animal Behaviour*, *34*(4), 1016–1025.

Fleeson, W. (2001). Toward a structure-and process-integrated view of personality: Traits as density distributions of states. *Journal of personality and social psychology*, *80*(6), 1011.

Fleeson, W. (2007). Situation□based contingencies underlying trait□content manifestation in behavior. *Journal of personality*, *75*(4), 825–862.

Fleeson, W., & Noftle, E. (2008). The End of the Person-Situation Debate: An Emerging Synthesis in the Answer to the Consistency Question. *Social and Personality Psychology Compass*, *2*(4), 1667–1684. https://doi.org/10.1111/j.1751-9004.2008.00122.x

Frankenhuis, W. E., & Nettle, D. (2020). Integration of plasticity research across disciplines. *Current Opinion in Behavioral Sciences*, *36*, 157–162.

Frankenhuis, W. E., & Tiokhin, L. (2018). Bridging evolutionary biology and developmental psychology: Toward an enduring theoretical infrastructure. *Child Development*, *89*(6), 2303–2306.

Funder, D. C. (1991). Global traits: A neo-Allportian approach to personality. *Psychological Science*, *2*(1), 31–39.

Funder, D. C., & Ozer, D. J. (1983). Behavior as a function of the situation. *Journal of Personality and Social Psychology*, *44*(1), 107–112. https://doi.org/10.1037/0022-3514.44.1.107

Furr, R. M., & Funder, D. C. (2004). Situational similarity and behavioral consistency: Subjective, objective, variable-centered, and person-centered approaches. *Journal of Research in Personality*, *38*(5), 421–447. https://doi.org/10.1016/j.jrp.2003.10.001

Garamszegi, L. Z., Markó, G., & Herczeg, G. (2012). A meta-analysis of correlated behaviours with implications for behavioural syndromes: Mean effect size, publication bias, phylogenetic effects and the role of mediator variables. *Evolutionary Ecology*, *26*(5), 1213–1235.

Geiser, C., Keller, B. T., Lockhart, G., Eid, M., Cole, D. A., & Koch, T. (2015). Distinguishing state variability from trait change in longitudinal data: The role of measurement (non) invariance in latent state-trait analyses. *Behavior Research Methods*, *47*(1), 172–203.

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological methods*, *19*(1), 72.

Geukes, K., van Zalk, M., & Back, M. D. (2018). Understanding personality development: An integrative state process model. *International Journal of Behavioral Development*, *42*(1), 43–51. https://doi.org/10.1177/0165025416677847

Giles, N., & Huntingford, F. A. (1984). Predation risk and inter-population variation in antipredator behaviour in the three-spined stickleback, Gasterosteus aculeatus L. *Animal behaviour*, *32*(1), 264–275.

Goddard, M. E., & Beilharz, R. G. (1984). A factor analysis of fearfulness in potential guide dogs. *Applied Animal Behaviour Science*, *12*(3), 253–265.

Goldberg, L. R. (1982). From Ace to Zombie: Some explorations in the language of personality. *Advances in personality assessment*, *1*, 203–234.

Goldberg, L. R. (1990). An alternative" description of personality": The big-five factor structure. *Journal of personality and social psychology*, *59*(6), 1216.

Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PloS one*, *12*(6), e0174035.

Goold, C., Vas, J., Olsen, C., & Newberry, R. C. (2016). Using network analysis to study behavioural phenotypes: An example using domestic dogs. *Royal Society Open Science*, *3*(10). https://doi.org/10.1098/rsos.160268

Gosling, S. D. (2001). From mice to men: What can we learn about personality from animal research? *Psychological bulletin*, *127*(1), 45.

Groothuis, T. G. G., & Carere, C. (2005). Avian personalities: Characterization and epigenesis. *Neuroscience and Biobehavioral Reviews*, *29*(1 SPEC. ISS.), 137–150. https://doi.org/10.1016/j.neubiorev.2004.06.010

Guloksuz, S., Pries, L. K., & Van Os, J. (2017). Application of network methods for understanding mental disorders: Pitfalls and promise. *Psychological medicine*, *47*(16), 2743–2752.

Gurven, M., Von Rueden, C., Massenkoff, M., Kaplan, H., & Lero Vie, M. (2013). How universal is the Big Five? Testing the five-factor model of personality variation among forager–farmers in the Bolivian Amazon. *Journal of personality and social psychology*, *104*(2), 354.

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, *10*(4), 255–282.

Guyon, H., Falissard, B., & Kop, J.-L. (2017). Modeling psychological attributes in psychology–an epistemological discussion: Network analysis vs. Latent variables. *Frontiers in psychology*, *8*, 798.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*(7302), 29.

Holtmann, B., Lagisz, M., & Nakagawa, S. (2017). Metabolic rates, and not hormone levels, are a likely mediator of between☐individual differences in behaviour: A meta☐analysis. *Functional Ecology*, *31*(3), 685–696.

Horstmann, K. T., Rauthmann, J. F., Sherman, R. A., & Ziegler, M. (2021). Distinguishing simple and residual consistencies in functionally equivalent and non-equivalent situations: Evidence from experimental and observational longitudinal data. *European Journal of Personality*, *35*(6), 833–860. https://doi.org/10.1177/08902070211014029

Huntingford, F., & Giles, N. (1987). Individual variation in anti‐predator responses in the three‐spined stickleback (Gasterosteus aculeatus L.). *Ethology*, *74*(3), 205–210.

Jackson, J. J., Wood, D., Bogg, T., Walton, K. E., Harms, P. D., & Roberts, B. W. (2010). What do conscientious people do? Development and validation of the Behavioral Indicators of Conscientiousness (BIC). *Journal of Research in Personality*, *44*(4), 501–511. https://doi.org/10.1016/j.jrp.2010.06.005

John, O. P., Goldberg, L. R., & Angleitner, A. (1984). Better than the alphabet: Taxonomies of personality-descriptive terms in English, Dutch, and German. *Personality psychology in Europe*, *1*.

Johnson, Wendy & Penke, Lars. (2014). Genetics of Social Behavior. In Gawronski, B & Bodenhausen, G (Hrsg.), *Theory and Explanation in Social Psychology*. The Guildford Press.

Jung, C. G. (1915). On psychological understanding. *The Journal of Abnormal Psychology*, *9*(6), 385–399. https://doi.org/10.1037/h0073967

Kain, J. S., Stokes, C., & de Bivort, B. L. (2012). Phototactic personality in fruit flies and its suppression by serotonin and white. *Proceedings of the National Academy of Sciences*, *109*(48), 19834–19839.

Klages, L. (1926). *Die Grundlagen der Charakterkunde*. J. A. Barth.

Kok, E. M. A., Burant, J. B., Dekinga, A., Manche, P., Saintonge, D., Piersma, T., & Mathot, K. J. (2019). Within-Individual Canalization Contributes to Age-Related Increases in Trait Repeatability: A Longitudinal Experiment in Red Knots. *The American Naturalist*, *194*(4), 455–469. https://doi.org/10.1086/704593

Koski, S. E. (2011). How to Measure Animal Personality and Why Does It Matter? Integrating the Psychological and Biological Approaches to Animal Personality. *From Genes to Animal Behavior: Social Structures, Personalities, Communication by Color*, 115–136. https://doi.org/10.1007/978-4-431-53892-9_5

Kotov, R., Krueger, R. F., Watson, D., Cicero, D. C., Conway, C. C., DeYoung, C. G., Eaton, N. R., Forbes, M. K., Hallquist, M. N., Latzman, R. D., Mullins-Sweatt, S. N., Ruggero, C. J., Simms, L. J., Waldman, I. D., Waszczuk, M. A., & Wright, A. G. C. (2021). The Hierarchical Taxonomy of Psychopathology (HiTOP): A Quantitative Nosology Based on Consensus of Evidence. *Annual Review of Clinical Psychology*, *17*(1), 83–108. https://doi.org/10.1146/annurev-clinpsy-081219-093304

Kralj-Fišer, S., & Schuett, W. (2014). Studying personality variation in invertebrates: Why bother? *Animal Behaviour*, *91*, 41–52. https://doi.org/10.1016/j.anbehav.2014.02.016

Laland, K., Matthews, B., & Feldman, M. W. (2016). An introduction to niche construction theory. *Evolutionary ecology*, *30*(2), 191–202.

Laskowski, K. L., & Bell, A. M. (2014). Strong personalities, not social niches, drive individual differences in social behaviours in sticklebacks. *Animal Behaviour*, *90*, 287–295.

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate behavioral research*, *39*(2), 329–358.

Lukaszewski, A. W., Gurven, M., von Rueden, C. R., & Schmitt, D. P. (2017). What explains personality covariation? A test of the socioecological complexity hypothesis. *Social Psychological and Personality Science*, *8*(8), 943–952.

MacDonald, K. (1983). Stability of individual differences in behavior in a litter of wolf cubs (Canis lupus). *Journal of Comparative Psychology*, *97*(2), 99.

Martin, J. S., Massen, J. J. M., Šlipogor, V., Bugnyar, T., Jaeggi, A. V., & Koski, S. E. (2019). The EGA+GNM framework: An integrative approach to modelling behavioural syndromes. *Methods in Ecology and Evolution*, *10*(2), 245–257. https://doi.org/10.1111/2041-210X.13100

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*(7), 1009–1020.

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological methods*, *23*(3), 412.

Međedović, J. (2018). What can human personality psychology learn from behavioral ecology? *Journal of Comparative Psychology*, *132*(4), 382.

Michelangeli, M., Goulet, C. T., Kang, H. S., Wong, B. B. M., & Chapple, D. G. (2018). Integrating thermal physiology within a syndrome: Locomotion, personality and habitat selection in an ectotherm. *Functional Ecology*, *32*(4), 970–981.

Mitchell, D. J., Beckmann, C., & Biro, P. A. (2021). Understanding the unexplained: The magnitude and correlates of individual differences in residual variance. *Ecology and Evolution*, *11*(12), 7201–7210.

Mitchell, D. J., & Biro, P. A. (2017). Is behavioural plasticity consistent across different environmental gradients and through time? *Proceedings of the Royal Society B: Biological Sciences*, *284*(1860), 20170893.

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, *2*(4), 201–218.

Molenaar, P. C. M., & Campbell, C. G. (2009). The New Person-Specific Paradigm in Psychology. *Current Directions in Psychological Science*, *18*(2), 112–117. https://doi.org/10.1111/j.1467-8721.2009.01619.x

Montiglio, P.-O., Ferrari, C., & Reale, D. (2013). Social niche specialization under constraints: Personality, social interactions and environmental heterogeneity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1618), 20120343.

Moretz, J. A., Martins, E. P., & Robison, B. D. (2007). Behavioral syndromes and the evolution of correlated behavior in zebrafish. *Behavioral ecology*, *18*(3), 556–562.

Mõttus, R. (2016a). Towards more rigorous personality trait–outcome research. *European Journal of Personality*, *30*(4), 292–303.

Mõttus, R. (2016b). Towards more rigorous personality trait–outcome research. *European Journal of Personality*, *30*(4), 292–303.

Mõttus, R., Briley, D. A., Zheng, A., Mann, F. D., Engelhardt, L. E., Tackett, J. L., Harden, K. P., & Tucker-Drob, E. M. (2019). Kids becoming less alike: A behavioral genetic analysis of developmental increases in personality variance from childhood to adolescence. *Journal of personality and social psychology*, *117*(3), 635.

Mõttus, R., Kandler, C., Bleidorn, W., Riemann, R., & McCrae, R. R. (2017). Personality traits below facets: The consensual validity, longitudinal stability, heritability, and utility of personality nuances. *Journal of Personality and Social Psychology*, *112*(3), 474.

Nakagawa, S., & Schielzeth, H. (2010). Repeatability for Gaussian and non☐Gaussian data: A practical guide for biologists. *Biological Reviews*, *85*(4), 935–956.

Negrão, N., & Schmidek, W. R. (1987). Individual differences in the behavior of rats (Rattus norvegicus). *Journal of Comparative Psychology*, *101*(2), 107.

Nettle, D. (2006). The evolution of personality variation in humans and other animals. *American Psychologist*, *61*(6), 622.

Nettle, D., & Penke, L. (2010). Personality: Bridging the literatures from human psychology and behavioural ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1560), 4043–4050. https://doi.org/10.1098/rstb.2010.0061

Niemelä, P. T., & Dingemanse, N. J. (2018). On the usage of single measurements in behavioural ecology research on individual differences. *Animal behaviour*, *145*, 99–105.

O'Dea, R. E., Noble, D. W. A., & Nakagawa, S. (2020). Unifying individual differences in personality, predictability and plasticity: A practical guide. *Methods in Ecology and Evolution*.

Oldham, J., & Morris, L. B. (2012). *The new personality self-portrait: Why you think, work, love and act the way you do*. Bantam.

Penke, L. (2010). Bridging the Gap Between Modern Evolutionary Psychology and the Study of Individual Differences. In D. M. Buss & P. H. Hawley (Hrsg.), *The Evolution of Personality and Individual Differences* (S. 243–279). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195372090.003.0009

Penke, L., Denissen, J. J. A., & Miller, G. F. (2007). The evolutionary genetics of personality. *European Journal of Personality: Published for the European Association of Personality Psychology*, *21*(5), 549–587.

Perals, D., Griffin, A. S., Bartomeus, I., & Sol, D. (2017). Revisiting the open-field test: What does it really tell us about animal personality? *Animal Behaviour*, *123*, 69–79.

Pigliucci, M. (2003). Phenotypic integration: Studying the ecology and evolution of complex phenotypes. *Ecology letters*, *6*(3), 265–272.

Polushina, T., Banerjee, N., Giddaluru, S., Bettella, F., Espeseth, T., Lundervold, A. J., Djurovic, S., Cichon, S., Hoffmann, P., & Nöthen, M. M. (2021). Identification of pleiotropy at the gene level between psychiatric disorders and related traits. *Translational psychiatry*, *11*(1), 1–9.

Roff, D. A., & Fairbairn, D. J. (2012). The evolution of trade☐offs under directional and correlational selection. *Evolution: International Journal of Organic Evolution*, *66*(8), 2461–2474.

Rolland, J.-P. (2002). *The cross-cultural generalizability of the Five-Factor model of personality.*

Rossier, J., Ouedraogo, A., Dahourou, D., Verardi, S., & Meyer de Stadelhofen, F. (2013). Personality and personality disorders in urban and rural Africa: Results from a field trial in Burkina Faso. *Frontiers in psychology*, *4*, 79.

Royauté, R., Buddle, C. M., & Vincent, C. (2014). Interpopulation variations in behavioral syndromes of a jumping spider from insecticide☐treated and insecticide☐free orchards. *Ethology*, *120*(2), 127–139.

Royauté, R., Hedrick, A., & Dochtermann, N. A. (2020). Behavioural syndromes shape evolutionary trajectories via conserved genetic architecture. *Proceedings of the Royal Society B*, *287*(1927), 20200183.

Royce, J. R., & Poley, W. (1975). Invariance of factors of mouse emotionality with changed experimental conditions. *Multivariate behavioral research*, *10*(4), 479–487.

Rushton, J. P., Brainerd, C. J., & Pressley, M. (1983). Behavioral development and construct validity: The principle of aggregation. *Psychological bulletin*, *94*(1), 18.

Saltz, J. B. (2019). Gene–environment correlation in humans: Lessons from psychology for quantitative genetics. *Journal of Heredity*, *110*(4), 455–466.

Sánchez-Tójar, A., Moiron, M., & Niemelä, P. T. (2022). Terminology use in animal personality research: A self-report questionnaire and a systematic review. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1968), 20212259. https://doi.org/10.1098/rspb.2021.2259

Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky, C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., & Araya☐Ajoy, Y. G. (2020). Robustness of linear mixed☐effects models to violations of distributional assumptions. *Methods in Ecology and Evolution*, *11*(9), 1141–1152.

Schwenk, K., & Wagner, G. P. (2001). Function and the evolution of phenotypic stability: Connecting pattern to process. *American Zoologist*, *41*(3), 552–563.

Shedler, J., Beck, A., Fonagy, P., Gabbard, G. O., Gunderson, J., Kernberg, O., Michels, R., & Westen, D. (2010). Personality disorders in DSM-5. *American Journal of Psychiatry*, *167*(9), 1026–1028.

Sih, A., Bell, A., & Johnson, J. C. (2004). Behavioral syndromes: An ecological and evolutionary overview. *Trends in Ecology and Evolution*, *19*(7), 372–378. https://doi.org/10.1016/j.tree.2004.04.009

Sih, A., Mathot, K. J., Moirón, M., Montiglio, P.-O., Wolf, M., & Dingemanse, N. J. (2015). Animal personality and state–behaviour feedbacks: A review and guide for empiricists. *Trends in ecology & evolution*, *30*(1), 50–60.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107.

Simmons, L. W. (2014). 25 years of behavioral ecology. *Behavioral Ecology*, *25*(1), 1–3. https://doi.org/10.1093/beheco/art108

Sinervo, B., & Svensson, E. (2002). Correlational selection and the evolution of genomic architecture. *Heredity*, *89*(5), 329–338.

Sinervo, B., Svensson, E., & Comendant, T. (2000). Density cycles and an offspring quantity and quality game driven by natural selection. *Nature*, *406*(6799), 985–988.

Sinn, D. L., & Moltschaniwskyj, N. A. (2005). Personality traits in dumpling squid (Euprymna tasmanica): Context-specific traits and their correlation with biological characteristics. *Journal of Comparative Psychology*, *119*(1), 99–110. https://doi.org/10.1037/0735-7036.119.1.99

Slater, P. J. B. (1981). Individual differences in animal behavior. *Perspectives in ethology*, 35–49.

Smaldino, P. E., Lukaszewski, A., von Rueden, C., & Gurven, M. (2019). Niche diversity can explain cross-cultural differences in personality structure. *Nature Human Behaviour*, *3*(12), 1276–1283.

Smith, J. M., & Price, G. R. (1973). The logic of animal conflict. *Nature*, *246*(5427), 15–18.

Soto, C. J., Kronauer, A., & Liang, J. K. (2016). Five-factor model of personality. *Encyclopedia of adulthood and aging*, *2*, 506–510. https://doi.org/10.1002/9781118521373.wbeaa014

Stamps, J., & Groothuis, T. G. G. (2010). The development of animal personality: Relevance, concepts and perspectives. *Biol Rev*, *85*. https://doi.org/10.1111/j.1469-185X.2009.00103.x

Stevenson-Hinde, J., & Zunz, M. (1978). Subjective assessment of individual rhesus monkeys. *Primates*, *19*(3), 473–482.

Steyer, R., Schmitt, M., & Eid, M. (1999). Latent state–trait theory and research in personality and individual differences. *European Journal of Personality*, *13*(5), 389–408.

Thalmayer, A. G., & Saucier, G. (2014). The questionnaire big six in 26 nations: Developing cross☐culturally applicable big six, big five and big two inventories. *European Journal of Personality*, *28*(5), 482–496.

Thorndike, E. L. (1924). Measurement of Intelligence. *Psychological Review*, *31*(3), 219.

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für tierpsychologie*, *20*(4), 410–433.

Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions: Congeneric and asymmetrical measurements. *Frontiers in psychology*, 7, 769.

Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Current directions in psychological science*, *9*(5), 160–164.

Turkheimer, E., Pettersson, E., & Horn, E. E. (2014). A Phenotypic Null Hypothesis for the Genetics of Personality. *Annual Review of Psychology*, *65*(1), 515–540. https://doi.org/10.1146/annurev-psych-113011-143752

Uher, J. (2008). Comparative personality research: Methodological approaches. *European Journal of Personality*, *22*(5), 427–455. https://doi.org/10.1002/per.680

Uher, J. (2011). Individual behavioral phenotypes: An integrative meta□theoretical framework. Why "behavioral syndromes" are not analogs of "personality". *Developmental Psychobiology*, *53*(6), 521–548.

Urszán, T. J., Garamszegi, L. Z., Nagy, G., Hettyey, A., Török, J., & Herczeg, G. (2018). Experience during development triggers between□individual variation in behavioural plasticity. *Journal of Animal Ecology*, *87*(5), 1264–1273.

Verbeek, M. E. M., Drent, P. J., & Wiepkema, P. R. (1994). Consistent individual differences in early exploratory behaviour of male great tits. *Animal Behaviour*, *48*(5), 1113–1121.

Vinkhuyzen, A. A. E., Pedersen, N. L., Yang, J., Lee, S. H., Magnusson, P. K. E., Iacono, W. G., McGue, M., Madden, P. A. F., Heath, A. C., & Luciano, M. (2012). Common SNPs explain some of the variation in the personality dimensions of neuroticism and extraversion. *Translational psychiatry*, *2*(4), e102–e102.

Wacker, J., & Smillie, L. D. (2015). Trait Extraversion and Dopamine Function. *Social and Personality Psychology Compass*, *9*(6), 225–238. https://doi.org/10.1111/spc3.12175

Weiss, A., Adams, M. J., Widdig, A., & Gerald, M. S. (2011). Rhesus Macaques (Macaca mulatta) as Living Fossils of Hominoid Personality and Subjective Well-Being. *Journal of Comparative Psychology*, *125*(1), 72–83. https://doi.org/10.1037/a0021187

Westneat, D. F., Wright, J., & Dingemanse, N. J. (2015). The biology hidden inside residual within□ individual phenotypic variation. *Biological Reviews*, *90*(3), 729–743.

Wilson, D. S. (1998). Adaptive individual differences within single populations. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *353*(1366), 199–205.

Wolf, M., & Weissing, F. J. (2010). An explanatory framework for adaptive personality differences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1560), 3959–3968.

Yamagata, S., Suzuki, A., Ando, J., Ono, Y., Kijima, N., Yoshimura, K., Ostendorf, F., Angleitner, A., Riemann, R., & Spinath, F. M. (2006). Is the genetic structure of human personality universal? A cross-cultural twin study from North America, Europe, and Asia. *Journal of personality and social psychology*, *90*(6), 987.

Appendix B: Manuscript II

Developing individual differences in primate behavior: The role of genes, environment and their interplay

Christoph J. von Borell [a, b *], Alexander Weiss [c,d], Lars Penke [a,b]

[a] Institute of Psychology, University of Goettingen, 37073 Göttingen, Germany

[b] Leibniz ScienceCampus Primate Cognition, 37077 Göttingen, Germany

[c] Department of Psychology, School of Philosophy, Psychology and Language Sciences, The University of Edinburgh, Edinburgh EH8 9JZ, United Kingdom

[d] Scottish Primate Research Group

*Corresponding author:

Christoph J. von Borell, Department of Biological Personality Psychology,

Georg Elias Müller Institute of Psychology, University of Goettingen,

37073 Göttingen, Germany

Phone: +49-(0)551-3920704

E-Mail: christoph.borell@psych.uni-goettingen.de

Abstract

As is the case for humans, it has long been thought that nonhuman primates can be described in terms of their personality. Scientific observations that support this view include the presence of individual differences in social behavior and that they are relatively stable throughout life. Consequently, individuals are constrained in their behavioral flexibility when dealing with various environmental challenges. Still, the variation among individuals during development suggests that the environment influences how primates behave. Research in fields including psychology, behavior genetics, and behavioral ecology have tried to identify the mechanisms responsible for this interplay of behavioral stability and change. In this review we integrate theories and findings from research on humans and nonhuman primates that highlight how and to what extent genetic and environmental contributions shape the development of social behavior. To do so we first provide an overview and define what is meant by mean level and rank-order change of behavior. We then review explanations of behavioral stability and change, focusing on the role of genetic effects, how environmental circumstances influence behavioral variation throughout development, and how genetic and environmental influences may interact to produce this variation. Finally, we point to future research directions that could help us to further understand the development of social behavior in primates from within a behavior genetics framework.

Introduction

Teasing, helping, playing, working, learning – within our circle of acquaintances, for many social behaviors, we can think of individuals that fall somewhere between one or the other extreme of variation in any given behavior. Apparently, social behavior and social relationships among humans are influenced by individual characteristics. Research from the last four decades has shown that this applies equally to our closest relatives, the nonhuman primates (henceforth "NHPs"). But how flexible are these individual characteristics? Where do they come from? And can they be changed? In this review we elaborate on the development of individual differences in behavior by comparing findings on humans and NHPs with a focus on the genetic and environmental forces that influence development.

In NHP personality research, the data underlying the quantification of individual differences typically stems either from questionnaires, completed by people with good knowledge of the individual animals, counted behavioral observations, or individuals' reactions to behavioral tests, where subjects encounter, for example, a setup containing novel objects or food items. Usually a variety of different behaviors are assessed, the correlations among behaviors are calculated and behaviors are grouped into summarizing dimensions using statistical techniques as factor analysis or principal component analysis. In humans, the investigation of such dimensions led to the formulation of the Five-Factor Model of human personality (Digman 1990), where differences among people can be summarized along the dimensions extraversion, agreeableness, conscientiousness, openness to experience, and neuroticism. The Five-Factor Model often serves as a reference point in NHP studies (see e.g. King and Figueredo 1997; Weiss et al. 2015) and analogues or variations of these factors have been found to a varying extent in different NHP species (Weiss 2017a).

The history of animal personality research and the different approaches used, whether by behavioral ecologists or comparative psychologists, have been reviewed elsewhere (Gosling 2001; Réale et al. 2007; Uher 2008; Koski 2011; Carter et al. 2013; Sih et al. 2015; Roche et al. 2016; Weiss 2017b). As such, we will not rehash this literature. Instead, we will focus on the development of behavioral variation among individuals. First, we will review the current knowledge about stability of behavioral differences on a

phenotypic level and then proceed to a more detailed overview of the genetic and environmental contributions to behavioral stability and change. We hereby will follow the broad conceptual separation common to research in behavior genetics. Hence by "genetic effects" we refer to behavioral variation due to differences in the sequence of the DNA of individuals and by "environmental effects" we refer to all other influences affecting behavioral variation that are not caused by variation in the individuals' DNA. Towards the end of our review we will also look at the interplay between genetic and environmental effects. The review will focus on findings from NHPs but will be complemented by findings from the human literature where appropriate, that is, if it provides additional insight.

Phenotypic stability over the lifetime

Do aggressive children grow up to be aggressive adults? To answer this and similar questions, we must distinguish between two types of behavioral stability or change. The first is an age-related metric called mean-level change, which refers to differences in the mean expression of a behavioral phenotype at different points in development. Mean-level change can be quantified with regression analysis where age (or different developmental stages, e.g., being an infant, juvenile, adult, etc.) is included as predictor of behavioral variation. Ideally, mean-level change is studied in a longitudinal design, with repeated measurements taken from the same individuals over time. The second is rank-order change, which is quantified by the magnitude of relative changes in behavior that occur among individuals within a population. It is independent of mean-level changes in absolute behavior. An example of a situation where there is little to no rank-order change would be if children who are highly aggressive relative to their age peers become adults who are highly aggressive relative to their age peers. Rank-order stability (or change) of behavior may be quantified by two techniques. The first involves conducting a simple correlation among behavioral measurements from two time points. The second involves computing the repeatability coefficient, which is an intraclass correlation that is based on multiple measures per individual and which describes the proportion of total behavioral variance due to differences between individuals (Boake 1989; Nakagawa and Schielzeth 2010). If the variance within individuals (between different measurements) is zero, then repeatability equals one. If the total behavioral variance is solely due to variation within individuals,

then repeatability equals zero. We illustrated the difference between mean-level and rank-order stability in Fig. 1.

Mean-Level Change

Knowledge of lifetime age effects on mean-level change in NHP personality stems especially from a study by King and colleagues (2008). They used cross-sectional data from chimpanzees that were divided into five age groups and found age-related differences in terms of lower extraversion and openness to experience scores, and higher agreeableness and conscientiousness scores, in older individuals. These results are corroborated by behavioral measurements from chimpanzees, where boldness and exploration tendency, which are related in their content to extraversion and openness, respectively, also appear to decline with age (Massen et al. 2013). Such a pattern could also be partly replicated in and transferred to orangutans by Weiss and King (2015), with the exception that in this species agreeableness is lower in older subjects. In common marmosets, females also tend to become less agreeable with increasing age, while both males and females become less inquisitive (Koski et al. 2017). The same pattern applies to older white-faced capuchin monkeys who are less agreeable and less open to new experiences as well (Manson and Perry 2013). So, although individuals are rather stable in their average behavioral propensities in relation to each other, age-related mean-level differences of behavior occur at the level of the population. Some age-related patterns seem to be similar across species (e.g., declines in openness / inquisitiveness / exploration tendency), while the development of agreeableness (indicating pro-social and tolerant behavior) differs among them. The reasons for developmental differences among species need to be clarified by future studies. Possible reasons for inter-species differences are the differing content and structure of the personality dimensions or varying selection pressures between species (Weiss and King 2015). In a sample of adult rhesus macaques (Brent et al. 2013), age was largely unrelated to personality dimensions, indicating that mean-level changes could be especially evident when changes over the lifetime or during early development are considered. Concerning the latter, strong changes in age-specific behavior have been reported that are tied to sex-specific life histories (Kulik et al. 2015a, b; von Borell et al. 2016).

In adult NHPs, the rank-order stability of behavioral differences ranges from being moderate (above r=0.3) to high (above r=0.5), and is statistically significant (e.g. King et al. 2008; Weiss et al. 2011; Brent et al. 2013; Weiss 2015). High levels of stability are found most often in studies that use ratings on questionnaires. Here, estimates of rank-order stability may be as high or higher than 0.7 (e.g., Stevenson-Hinde and Zunz 1978; King et al. 2008; Weiss et al. 2011). These estimates reflect the relative stability of *average* behavior of individuals, that is, the consistency of displaying certain behavioral phenotypes accumulated across situations (Weiss et al. 2009). If rank-order stability is calculated as repeatability of behavioral measurements, the resulting repeatability coefficient is typically lower than in studies using questionnaire ratings (e.g. Brent et al. 2013; Neumann et al. 2013; von Borell et al. 2016), aligning closer to the meta-analytical mean repeatability of 0.37 measured across species (Bell et al. 2009). It must be noted though that differences in repeatability among non-aggregated behavioral measurements and aggregated questionnaire ratings could occur because averaging single ratings into broader dimensions, that is, into personality "factors", "domains", "dimensions", or "components", contributes to the stability of these measures (Rushton et al. 1983). During early ontogeny, the stability of individual differences is typically lower than in adults (von Borell et al. 2016) and may show substantial variation from year to year, which may in turn differ across personality domains (Stevenson-Hinde et al. 1980).

What do our measurements tell us about stability? And what do they not tell us?

The studies presented so far used questionnaire ratings or counted behavioral observations to assess the personalities of the individuals under study. They showed patterns of mean-level change in behavior and rank-order stability of individual differences in behavior that predominantly reflect variation on a year-wise or season-wise timescale. However, these approaches may not be sensitive to short-termed effects of the environment. As indicated above, questionnaire ratings accumulate impressions of an animal's behavior

across situations and therefore do not capture short-term interactions of behavior with environmental fluctuations. Some of the studies also rely on animals kept in captivity (e.g., living in zoos, as in King et al. 2008), which may limit the naturally occurring environmental variation for some species.

One possible means by which the influence of the environment on behavior could be tested is by continuously sampling behavioral observations in free ranging animals (von Borell et al. 2016). Yet, the fallacy of behavioral sampling is that observations, for example single incidents of displaying aggressive behavior, are typically also aggregated over time to form a reliable estimate of individual propensities. Otherwise, rare coincidences, like a generally unaggressive individual showing a sign of aggression, could lead to unwarranted conclusions about a general behavioral tendency. Because naturally occurring observations of certain behaviors may be scarce, aggregation operates usually on relatively large time scales (e.g., year-wise or season-wise). Such aggregation limits the possibility of analyzing behavioral plasticity in response to the environment to long-term fluctuations, stable population differences, or permanent changes within populations (such environmental effects will be discussed in the following section). Whether there are developmental influences on short-term plasticity (i.e., reaction norms; Dingemanse et al. 2010) is thus often not assessed. This is despite the fact that it might be hypothesized that NHPs become, for example, less flexible in their behavior with increasing age. Examples from other species show that individuals may vary in their seasonal plasticity, that they are repeatable in such plasticity (i.e., temporally consistent in their rank-order of shown plasticity) and that the mean plasticity across individuals may decrease with age (e.g., in great tits; Araya-Ajoy and Dingemanse 2017). These findings of differences in plasticity are likely due to frequency-dependent costs or benefits leading to individually different behavioral strategies. Furthermore, such costs or benefits are likely to change with experience, leading to mean level changes in plasticity during ontogenic development (Wolf et al. 2008). The question of age-related variability in behavioral plasticity appears to be somewhat of a blind spot in the study of NHP behavioral development. To address this question requires studies that obtain repeated measurements of behavior-situation interactions within and across time intervals or that can calculate the effect of age on behavioral reaction norms in cross-sectional data. One way to gather these kinds of data is by means of behavioral

tests that involve simulating situations that an animal may encounter in the wild (e.g., encountering a novel environment or object, confrontation with the vocalization of a predator). For NHPs in captivity behavioral tests have been developed to assess behavioral variation among individuals (e.g., Uher et al. 2013; Staes et al. 2016). If such behavioral tests are conducted with environmental variation or transferred to the natural habitats of NHPs, this approach allows for a controlled collection of data that may be linked to short-term environmental fluctuations. For example, tests of social facilitation that compare behavioral responses to novelty when individuals are alone to when they are in a social context show short-term environmental effects on behavior (reviewed in Forss et al. 2017). In common marmosets, the latency to eat novel food is reduced in a social context, but only in juveniles, suggesting that individual age affects the strength of social facilitation (Yamamoto and Lopes 2004). Following these results, behavioral reaction norms of neophobia or exploration tendency with varying social contexts could be further tested in a longitudinal setting to assess the degree to which individual differences in reaction norms are stable throughout development, i.e., their rank-order stability. There are also examples of behavioral tests conducted with NHPs in the wild (e.g., playback experiments in Neumann et al. 2013; novel-object and novel-food tests in Arnaud et al. 2017). These could be paired with environmental information (e.g., current group composition, time elapsed since among-group conflict, etc.) to form behavioral reaction norms and tested for hypothesized age effects, preferably in a longitudinal design. Other possibilities would be to use data from continuous observations in a non-aggregated way or aggregating observations according to relatively short-term environmental fluctuations and analyze them via linear mixed effects models that can account for zero-inflated observations in the case of rarely observed behaviors (Zuur et al. 2009; Dingemanse and Dochtermann 2013; Brooks et al. 2017). Such an approach would be informative about relationships between behaviors, between individuals, (correlated) changes in behavior within individuals, and whether the interaction among behavior and environmental factors (plasticity) changes with age (Dingemanse and Dochtermann 2013). For a "how-to" example of using the full potential of linear mixed models when analyzing behavioral observations of NHPs see Martin and Suarez (2017).

What do we know from humans?

Findings from research on human personality development are largely consistent with findings from NHPs. In terms of rank-order stability, humans become more stable throughout their lives, developing from moderate stability (approx. r=0.35) in behavioral differences during childhood to high stability (approx. r=0.70) during late adulthood (Roberts and DelVecchio 2000; Terracciano et al. 2006). Mean-level changes occur primarily during early adulthood, a time often marked by major changes in an individual's environment and increased control over life-history decisions: After a period of decreased psychological "maturity" during early puberty (Denissen et al. 2013), humans typically develop towards a more mature and functional personality in that they become more agreeable, conscientious and show more emotional stability (Roberts et al. 2006; Donnellan et al. 2007). However, they also tend to become less flexible (Roberts et al. 2002).

Determinants of Plasticity and Stability in Behavior

Now that we know that behavioral variation among individuals is not fixed and that rank-order and mean-level changes occur in particular during childhood, adolescence, and young adulthood, the question remains how these changes can be explained. We propose to approach questions about behavioral stability and change using a behavior genetics framework, because it helps us to disentangle whether and how behavioral development is caused by environmental influences, genetic effects, or their interplay.

Genetic Effects on Behavioral Development

The rationale behind genetic effects on behavior is that variation in DNA sequences among individuals will lead to variation in their behavioral propensities. The extent to which genes influence a behavioral phenotype is measured with a population statistic "heritability". Heritability (or h^2) is the ratio of genetically influenced variance in a trait to the total variance of the trait in a population (Plomin et al. 2012; Johnson 2014). Heritability may also be calculated as the ratio of genetically influenced variance to the repeatable variance (as this "error-free" variance poses an upper limit to the heritability; Adams et al. 2012). A trait's

heritability may reflect additive genetic effects whereby the effects of variants of genes (polymorphisms) independently add up to shape the trait into a specific direction. This is known as narrow-sense heritability. A trait's heritability may also reflect non-additive genetic effects whereby the interactions among different gene variants affect the expression of the trait. An example of this would be a dominant genetic variant (allele) that suppresses the effect of a recessive genetic variant at the same or different loci. The combined influence of additive and non-additive genetic variance is referred to as broad-sense heritability, which is denoted H^2.

To provide a general impression of how heritable personality traits are in NHPs, we calculated the median and range of published estimates of narrow-sense heritability across NHP species and studies (see Tables S1, S2 in the supplement). For personality factors we calculated a median heritability of $h^2=0.25$ and a range from 0.00 to 0.63 (based on the studies from Weiss et al. 2000; Fairbanks et al. 2004; Adams et al. 2012; Brent et al. 2013; Johnson et al. 2015; Latzman et al. 2015; Staes et al. 2016; Wilson et al. 2017; Inoue-Murayama et al. 2018). The heritability of single behaviors appears to be very similar, with a median $h^2=0.25$ and range of 0.11 to 0.91 (based on studies by Rogers et al. 2008; Fawcett et al. 2014; Hopkins et al. 2014, 2015; Johnson et al. 2015; Watson et al. 2015). Non-additive genetic effects may contribute a significant proportion to genetically influenced variance, leading to higher broad-sense heritability estimates (H^2). Based on a study on orangutans we calculated a median H^2 of 0.69 (Adams et al. 2012). Published estimates of broad-sense heritability are, however, an exception, as this requires extended study designs including twins or a large number of full- and half-siblings (ibid.). Unfortunately for a developmental perspective, we do not know of longitudinal studies that published heritability estimates for a birth cohort across time. Nor do we know of cross-sectional estimates of heritability along different developmental stages. Hence, we cannot say whether the heritability estimates of personality traits, and thus influences relating to environmental factors, increase or decrease throughout development.

In humans the average heritability estimated from meta-analyses is a little higher than in NHPs, accounting for about 40% of variation (Turkheimer et al. 2014; Vukasović and Bratko 2015). Interestingly, estimates

coming from family and adoption studies, that include only additive genetic effects, have an average effect size of 0.22 (Vukasović and Bratko 2015), which is close to the median effect size we calculated for narrow-sense heritability in NHPs. This percentage may rise to about 50% when only data from twin studies is considered (van den Berg et al. 2004; Vukasović and Bratko 2015) as these estimates reflect the broad-sense heritability. From a developmental perspective, we know that the heritability of personality tends to decrease with increasing age, dropping from roughly 75% during infancy and early childhood down to the above-mentioned estimate of 40% in later adulthood (Briley and Tucker-Drob 2017). Thus, in the period after birth, individual differences in behavior are largely influenced by genetic effects, with the role of environmental effects increasing with age.

The increasing role of the environment is also reflected in its contribution to the increase in the rank-order stability of personality (from r=0.35 in infancy to about r=0.70 in adults; see above), which can be explained by genetic or environmental influences. Here twin studies find that the genetic contribution remains at a steady 35% during the lifespan, while the environmental contribution increases to account for an additional 35% of rank-order stability during development. This means that the stable proportion of behavioral variation is almost entirely genetically influenced during infancy, but that the post-infancy stability increase is almost entirely influenced by environmental factors (Tucker-Drob and Briley 2019).

Environmental Effects

Given the heritability estimates above, we can expect that environmental effects may contribute to over 50% of behavioral variation in NHPs and about 50% in humans, varying with the age of the individual. An important goal of personality and developmental studies across disciplines has been to identify environmental factors that are capable of altering or shaping behavioral differences among individuals. Here we review two broad categories of well-studied environmental factors that influence developing behavioral differences: stressful life experiences and the influence of maternal care and rearing conditions.

Stressful Life Experiences

Environmental stressors influence behavioral development during prenatal or very early life stages. For example, low food availability is linked to higher prenatal maternal stress in Assamese macaques, which leads to increased growth, but decreased motor skill acquisition and reduced immune function in their offspring (Berghänel et al. 2016). Although this evidence is circumstantial, life-history trade-offs such as these may extend to the development of individual differences in related behavioral traits, for example a trade-off between playing and growth (Berghänel et al. 2015). Fertility is also affected by low-quality early environments with individual differences being linked to drought years in baboons (Lea et al. 2015). Next to the quality of the environment, effects of the dominance hierarchy have been documented as a lasting stressor in NHP development. In chimpanzees, for example, maternal rank during pregnancy is not only related to the stress response of the mother, but also to the stress response of her dependent offspring, and especially males thereof (Murray et al. 2018). A relationship between maternal or individual rank and behavioral differences, and especially those relating to aggressive and fearful/bold behavior, has been shown for NHPs of different ages (e.g., French 1981; Bolig et al. 1992; Brent et al. 2013; von Borell et al. 2016). In an experimental manipulation, Kohn and colleagues (2016) showed that climbing up the dominance hierarchy was causally related to changes in social approachability and boldness. We can thus expect changes in the dominance hierarchy as a possible source of environmentally induced variation in personality development. Related evidence stems from a case of severe and selective tuberculosis infection in wild baboons, where the more aggressive individuals of a troop died at once, because they ate from a neighboring troop's food resource that was infected. These deaths led to an overall more tolerant social style in the troop. While dominance interactions were concentrated among closely ranked individuals, high-ranking individuals were more tolerant of very low-ranking individuals. The latter finding was related to a disproportionally high number of reversals in the direction of dominance among individuals far apart in rank (Sapolsky and Share 2004). This is in line with the argument that high-ranking individuals can typically afford aggressive or displacing behavior due to agonistic support from other individuals (Silk 2002), which was apparently less the case in the newly stratified troop of baboons after the epidemic infection.

Although the quality of the natural environment and dominance hierarchies in social groups affect behavioral differences from early life on, new challenges arise around the time of maturation that drive behavioral variation. A prominent example in NHPs is the migration from the natal group to a new group (natal dispersal). Migration is typically accompanied by increases in mortality or injury rates, decreases in access to resources, and social costs, i.e., the loss of social ties or rank (Dittus 1979; Weiß et al. 2016). Following migration, male rhesus macaques show more fearful and less physically aggressive behavior than before (von Borell et al. 2016), which is consistent with findings from captive pigtailed macaques, where individuals that are new to a facility are more cautious (Sussman et al. 2014). Migration may also trigger rank-order changes in behavior, possibly reflecting different reactions or strategies following migration. In the study of von Borell et al. (2016) this was reflected in very low or even negative correlations among fearful behaviors measured in the year before and after migration, despite their overall lifetime repeatability. In female rhesus macaques, the birth of the first infant is a similar developmental milestone and is marked by a decreased frequency of initiating social contacts outside of maternal kin (von Borell et al. 2016).

Maternal Influences and Rearing

Parental care and the quality of mother-offspring interactions are also known to affect the development of individual differences in NHP behavior. Here we highlight some findings in this literature. Interested readers are encouraged to refer to a detailed review of this literature in this topical collection (Maestripieri 2018).

Differences in maternal style are typically described along the two dimensions protectiveness and rejection, but may vary a little between NHP species, that is, maternal behaviors may also load on three different factors in a factor analysis (De Lathouwers and Van Elsacker 2004). Protectiveness and rejection have been linked to individual differences in behavior across various age-stages in NHP development. For example, in an observational study of Japanese macaques, infants of highly protective mothers showed lower levels of exploratory behavior and interacted less with their group members. On the other hand, infants of mothers who rejected them interacted more than average with other group members. These effects diminished, however, over the course of development and were present mostly during early infancy (Bardi and Huffman

2002). A stable effect of maternal style was reported by Bardi and colleagues (2015) who found that juvenile baboons that experienced more stress-related interactions with their mother during early life showed higher locomotor activity and cortisol levels during a stress test than individuals that experienced more affiliative mother-offspring interactions.

Such effects of parental care or mother-offspring interaction were further supported by experimental studies. An effect of maternal protectiveness on offspring caution was shown in vervet monkeys (Fairbanks and McGuire 1993). In this study maternal protectiveness was experimentally increased by introducing new males to some housing groups. Infants and juveniles of mothers from the "protective" condition showed higher latencies to approach a novel object, indicating increased caution. Approach latencies were highly correlated among mothers and infants but not among mothers and juveniles. These results indicate that a mixture of environmental and genetic effects contributed to the development of behavioral differences. Maestripieri and colleagues (2006) could not find an effect of maternal protectiveness on offspring behavior in rhesus macaques, but they did find that higher maternal rejection led to more solitary play in offspring. This effect did not differ between mother-reared and cross-fostered individuals, ruling out the possibility that this observation is simply driven by genetic similarity between mothers and their offspring.

A special case of maternal influence on behavioral differences is maternal deprivation or the disruption of maternal care. Rhesus macaques that spent their first year of life in total isolation showed hardly any positive social responses or activities afterwards and were also consistently fearful. Individuals who spent shorter periods of time in isolation showed a behavioral pattern similar to that of monkeys who spent a year in isolation, followed by highly individualized (adequate and non-adequate) adaptations to social situations, presumably based on inherited individual differences and unique learning experiences (Harlow et al. 1965). Similar differences in the social response to short periods of isolation have been documented in free-ranging rhesus macaque infants (Berman et al. 1994). Here, increased short-term separations of mothers and their infants, which occurred when the mothers resumed mating, led to increased distress in the infants. Like the captive infants, described by Harlow and colleagues (1965), who were isolated for

short periods, the free-ranging infants developed differing social responses to and after the separation events. Specifically, some infants reacted with social withdrawal and decrease of social play and others rather increased their social behavior like grooming. Differential responses to maternal separation or maternal style, whether marked by decreased or increased social behavior, have been linked with genetically inherited differences in stress responsivity (Clarke and Boinski 1995; Suomi 2004). Further studies of maternal separation in captivity, typically on hand-raised and later on peer-reared individuals, suggest temporally consistent increases in anxious, shy, and impulsive behavior in comparison with their mother-reared counterparts. These behavioral differences may extend to neglectful or abusive maternal behavior, when peer-reared females become mothers themselves (reviewed in Soumi 1997). More recent studies, albeit in a different species, show mixed results: while nursery-reared chimpanzees were reported to be less agreeable and more extraverted than their mother-reared counterparts (Latzman et al. 2015), a similar study of chimpanzees found no such differences between these groups (Martin 2005).

The effects of differential care appear to extend to scenarios were the intensity of human care varies. Young chimpanzees who experienced enhanced responsive care were less distressed and showed less disorganized attachment than chimpanzees who only received a minimal standard of care from human caregivers (van IJzendoorn et al. 2008). In addition to maternal style, maternal separation, and the amount of care, the time infants spend with conspecifics seems to affect personality development. For example, chimpanzees who as infants spent less time with conspecifics were rated as being less extraverted later in life than individuals who spent more time with conspecifics (Freeman et al. 2016).

Issues of causality

From a behavior genetics standpoint, non-experimental studies and non-genetically-informed quasi-experimental studies cannot establish causal relationships between environmental and behavioral variation. Although environmental effects can be separated in a controlled randomized experiment (at the cost of decreased ecological validity), all other behavior-environment correlations are likely influenced by genetic variation. As Johnson (2014) put it:

"The situation and the individual's environmental history may set the stage and limit the range of choice of action, but the individual's genotype is involved both in the actions taken and the individual's presence in this situation in the first place. We cannot understand development without taking this into consideration."

Among the findings on stressful life events or rearing experience reviewed above, experimentally separated environmental effects rely largely on captive NHPs, while in studies conducted in the wild, environmental and genetic effects can be confounded. There are several mechanisms of such confounding. Prominent examples include gene-environment correlations (rGE) and gene-environment interactions (G x E), both of which will be discussed below. The main message at this point is, that a neglect of genetic information can lead to premature causal interpretations of the role the environment may play in behavioral development (Briley et al. 2018). For example, the association between early adversity and a faster life-history strategy that has been reported in NHPs, has received theoretical and empirical support from the human literature as well, leading, for example, to earlier puberty and marriage (see reviews by Belsky 2012; Del Guidice 2014). However, findings of life-history embedded behavioral differences related to early adversity did not hold up in a study design that included information of genetic relatedness based on pedigrees to control for genetic confounding. Mendle and colleagues (2009) found that the association among father absence and timing of first intercourse in humans was best explained by genetic risk factors that correlate both with father absence and early sexual activity, diminishing the role of the mere *experience* of an absent father. Likewise, decisions involving changes in the social environment, such as NHP dispersal, are known to carry a genetic component (Trefilov et al. 2000; Krawczak et al. 2005) that could also be correlated to behavioral differences. Also, relationships between rank and behavior may partly be affected by feedback processes entailing a genetic component, for example the interplays of aggressive behavior, which has a heritable component, and changes in the dominance hierarchy in male NHPs (Koyama 1970; Bernstein 1976). In humans, some studies on personality development try to test whether environmental effects are causal by including a control group. Examples can be found in studies on personality development during periods of spatial and social transformation in human adolescents or young adults: events like a high-school student

exchange (Hutteman et al. 2015), studying abroad as college student (Zimmermann and Neyer 2013), graduation from high school (Bleidorn 2012), or forming a partner relationship (Neyer and Lehnart 2007) mostly trigger a development towards personality maturation compared to the control group, i.e., increases in conscientiousness, agreeableness and self-esteem, and a decrease in neuroticism. Going abroad was also related to increases in openness to new experiences. The inclusion of a control group is certainly an improvement over not including a control group, as it can be the case in related studies of NHP migration in the wild, where it is often difficult to gather a control group with similar characteristics and a similar sample size as the migrating individuals. Yet, in naturally occurring control group designs, such as the above-described human studies, the decision of whether to participate and the behavioral differences among individuals of the control and quasi-experimental groups may be influenced by common genetic effects. Even if both groups have been matched to be similar in their behavioral characteristics prior to the environmental change, this change may only activate or amplify a genetic predisposition of a behavioral tendency, for example, being open to new experience that was already entailed in the decision of participating in this event.

In the human literature, the impact of individuals' genetic background on behavior or (life-history) decisions (e.g., student exchange, marriage, etc.) led to the "first law of behavior genetics" that all traits are heritable (Turkheimer 2000). It follows that behavior-environment correlations cannot be interpreted as prima facie evidence of a *causal* environmental influence without considering that such associations are probably genetically mediated (Johnson et al. 2011; Johnson and Penke 2014; Turkheimer et al. 2014). Accordingly, calls for genetically informed designs in the study of behavior-environment associations have been pointed out in primatology (e.g. Adams 2014; Brent and Melin 2014) and psychology (Turkheimer and Harden 2014), that could control for a genetic basis of differences in the environment that individuals experience. For example, studies looking at the effects of migration on behavioral differences among individuals could control for the possibility that both share a common genetic basis. Briley and colleagues (2018) reviewed techniques that are capable of tackling questions of causality in longitudinal, and even cross-sectional, genetically informative data (i.e., data where behavioral outcomes and measurements of the environment

are paired with information about relatedness or molecular genetic similarity among individuals). For example, in a quantitative genetic design, direction-of-causation modeling (DOC modeling) can be used to estimate the plausibility of a causal direction among an environmental and a behavioral measure. This approach involves comparing the proportion of variance attributable to genetic, shared, and nonshared environmental effects in the possible cause and outcome. If, for example, differences in maternal style have a large genetic component and causally explain behavioral differences among children, then a genetic component should be represented in the children's behavioral differences as well. Comparing the fit of different models with alternative directions of causality can help to assess the likelihood of a hypothesized cause-outcome-relationship (for details see Briley et al. 2018). In human female twins, DOC modeling showed that parental behavior was more likely the cause of psychological distress than psychological distress being the cause of parental behavior (i.e., the model specifying a causal relationship from parental behavior to distress had a better fit than the other way around; Gillespie et al. 2003).

Gene-environment interplay

As pointed out above, in observational studies, whenever a complex interplay among genes and the environment is present during development, separating the environmental and genetic sources of variance can be difficult (but still see Briley et al. 2018). In the case of gene-environment correlations (rGE), individuals evoke, pick, or create environmental experiences based on genetically influenced needs or preferences, or grow up in an environment that is influenced by genes they share with their parents (see, e.g., Scarr and McCartney 1983; Bleidorn et al. 2014; Weiss 2017b). Another possibility is that the impact of environmental experiences differs depending on individuals' genetic backgrounds (e.g., a genetic risk or vulnerability; Moffitt 2005), which is termed gene-environment interaction (G x E). While heritability estimates tell us that the biological underpinnings of behavior cannot be ignored in developmental studies, they are less useful in helping us to understand the developmental mechanisms or processes behind emerging behavioral differences, as variance is here partitioned into being genetic or environmental, and so does not account for gene-environment interplay (Plomin and Bergeman 1991).

In some species, it is possible to conduct controlled experiments on developmental psychobiology that allow for a separation of genetic and environmental effects (e.g., by breeding genetically identical individuals in identical conditions; Kain et al. 2012; Bierbach et al. 2017), but ethical and practical reasons mostly prevent scientists from applying these methods to humans or NHPs (Turkheimer 2000; but see experimental manipulations of rearing conditions presented above). Yet, there is no need for primatologists or psychologists to stop searching for the causes of development. Although we may not be able to causally reconstruct complex developmental pathways, we can test how genes and the environment correlate and interact in specific scenarios and how likely they are to shape behavioral development within the limits of such scenarios.

An example of NHP rGEs is the above-cited genetic influence on dispersal where genetic variation leads to different ages of migration from the natal group, that is, the encounter of a novel environment (Trefilov et al. 2000). Correlations among genes (or genetically influenced traits) and the environment are often referred to as "niche picking" or "niche specialization" (Johnson et al. 2009; Penke 2010; Stamps and Groothuis 2010; for evolutionary and mathematical formalization, see Montiglio et al. 2013). If we consider a developmental pathway where having more of some trait leads to a higher propensity to seek out a specific environment, which in turn affects the manifestation of that trait, then cross-sectional studies cannot distinguish between such bidirectional influences of genetic background and the environment (Kandler et al. 2012). If not explicitly modeled, the variation due to rGE will be confounded with genetic variance, although an environmental influence is entailed as well (Bleidorn et al. 2014). Genetically informed longitudinal studies, however, make it possible to test instantiations of rGE. In humans, Kandler and colleagues (2012) showed that genetic effects on personality traits, such as neuroticism or agreeableness, can explain variation in the likelihood of experiencing negative life events and that negative life events, in turn, have a (small) effect on personality development.

G x E effects on personality development can be detected by quantitative or molecular genetics methods. Quantitative genetic studies test whether differences in a phenotype between individuals are associated with information on their genetic relatedness (for example based on known pedigrees), while molecular genetic studies try to associate differences in a phenotype with a specific pattern of variation in DNA sequence among individuals. In behavioral genetic research, the latter's emphasis is on trying to find associations between genetic variants at specific genetic loci and behavioral traits (candidate gene association study) or trying to associate a large number of variants that are spread across the genome with a behavioral trait (genome-wide association study, GWAS). In a quantitative genetics framework, Latzman and colleagues (2015) have shown that heritability estimates of personality dimensions vary among mother- and nursery-reared chimpanzees. Specifically, they found lower heritability estimates in nursery-reared individuals indicating that their atypical environmental circumstances at an early age led to a higher proportion of environmentally influenced behavioral variation among their traits. Results from humans also support interaction effects of rearing quality and genes. For example, Krueger and colleagues (2008) showed that the genetic influence on adolescent personality varied with the levels of regard they received from their parents. In particular, low levels of regard were associated with an increased environmental contribution to phenotypic variance. On a molecular level, many NHP studies have examined the interplay of environmental variation and candidate genes in their contribution to behavioral differences. These studies analyzed for example polymorphisms in genes such as 5-HTTLPR (Barr et al. 2004; Madrid et al. 2018), MAOA (Newman et al. 2005), and COMT (Gutleb et al. 2017), which often, but not exclusively, were reported to interact with differences in rearing condition (for a review see Rogers 2018).

In the molecular genetics area, studies of NHPs and humans used to be closely linked and shared a desire to identify the genetic underpinnings of behavioral or pathological variation by testing the effects of candidate genes (see, e.g., Caspi et al. 2002, 2003 on G x E in humans, including MAOA and 5-HTTLPR variation affecting violence and depression, respectively). However, meta-analyses and recent studies in humans that use samples that are several magnitudes larger in size and extensive genome-wide genetic information led to the conclusion that complex behavioral traits are unlikely to be substantially influenced

by single genes (Munafò and Flint 2004; Plomin and von Stumm 2018; Sallis et al. 2018). That does not mean that genetic polymorphisms in single genes do not matter, but that their effects are usually too small to be detected with the sample sizes of earlier studies, and this is especially the case when they are modeled in interactions with environmental gradients. Reviews of human candidate gene studies show that many associations cannot be replicated across studies and in meta-analyses, and that the effect sizes of statistically significant associations in earlier studies were often inflated (e.g. Sanchez-Roige et al. 2018). These findings led researchers to conclude that the literature on associations among common variants in candidate genes and behavior, for both main effects and G x E interactions, is awash with false positive results (Sallis et al. 2018). Genome-wide association studies that explore associations of common genetic variants and behavior throughout the whole genome show that a large number of genetic variants (single nucleotide polymorphisms; SNPs) contribute to the heritability of complex traits, however with small effect sizes. Replicated SNPs typically explain less than 0.1% of the phenotypic variance (Munafò et al. 2014; Sallis et al. 2018). While many SNPs reported in candidate gene studies did not replicate in sufficiently powered GWAS (e.g., Chabris et al. 2012), many variants that met genome-wide significance levels that have been identified in GWAS could be replicated in large independent samples (> 100,000 individuals; e.g., Okbay et al. 2016). These variants are spread broadly across the genome, including intragenic regions that do not code for proteins (Boyle et al. 2017; Sanchez-Roige et al. 2018). Additionally, extended study designs show that rare genetic variants that are not tagged in GWAS can contribute to individual variation in complex traits (Hill et al. 2018). While these findings and conclusions stem from human studies, they are likely to apply to NHP studies as well (Munafò et al. 2014). That is not to say that all statistically significant results stemming from NHP candidate-gene or GxE studies are false positives. Some gene-behavior associations have replicated across populations, species, and behavioral measures (reviewed in Weiss 2017a; Rogers 2018). For example, variants in the arginine vasopressin receptor 1A gene (AVPR1A) appear to replicate across different samples of chimpanzees (Anestis et al. 2014; Hopkins et al. 2014; Staes et al. 2015; Wilson et al. 2017), bonobos (Staes et al. 2016) and common marmosets (Inoue-Murayama et al. 2018). However, the combination of small sample sizes and relatively large effects of reported genetic variants is similar to the early wave of human studies in the field of behavior genetics. It is thus probably worth retaining one's

skepticism about this literature. Reported effect sizes of replicated genetic variants in NHPs (e.g. given in Staes et al. 2015 and Wilson et al. 2017 for AVPR1A) are several magnitudes larger than most of the extensively studied candidate-gene variants and GWAS results in humans (see Sanchez-Roige et al. 2018 for a review). It is possible that the development and the social influences on behavioral variation among humans are more complex and thus less influenced by single genetic variants. Also, studies on captive NHPs provide a more restricted and controlled environment (e.g., controlled diet, less habitat variation), which might lead to stronger genetic effects. A recent study on the effects of variants in OXTR and AVP receptor genes (AVPR1A, AVPR1B) on behavior in rhesus macaques, however, failed to replicate previous results and showed only very small effects of the 12 SNPs that were examined (Madlon-Kay et al. 2018). Alongside the emerging consistency of findings that single genetic variants have only small effects on complex traits, Madlon-Kay and colleagues (2018) discuss other methodological difficulties, including missing control of genetic relatedness within the population and/or missing adjustment of p-values, that raise doubt about earlier positive results.

A promising avenue for matching smaller sample sizes with genetic information appears to be the use of polygenic scores, where genetic variants accounting for small effects are weighted and summed, creating a score for each subject that is a more powerful estimator of behavioral differences. Given a robust knowledge of genetic variants that contribute to behavioral differences in a species, polygenic scores can help relatively small samples to reach sufficient power to detect molecular genetic effects on behavior and be paired with environmental measures to assess G x E (Plomin and von Stumm 2018). For example, a polygenic score that predicts 10% of the variance in a trait only needs a sample size of 60 individuals to detect its effect with 80% power (ibid.). The problem for NHP studies is that, depending on the species, it might be impossible to gather a sufficiently large initial sample to identify genetic variants that are worth including in a polygenic score in the first place.

In the concluding lines of this section we want to provide a glimpse into the emerging field of epigenetics. Epigenetics refers to processes whereby environmental signals affect genetic variation by mechanisms such

as DNA methylation or histone modification. Briefly, these environmentally induced mechanisms can lead to individual differences in gene transcription and expression, which can result in behavioral differences (Kaminsky et al. 2008). In baboons, for example, Runcie and colleagues (2013) found that different aspects of the social environment and social behavior (social connectedness, group size, and maternal dominance rank) interacted with the genotype by means of differences in gene expression along these environmental or behavioral gradients. This suggests that social behaviors, like grooming, are not only influenced by genetic variation, but also influence genetic variation. From an ontogenetic perspective, this means that genes are not destiny for the development of personality, but rather that the environment can alter the genetic tracks individuals are set on. The precise way in which epigenetic mechanisms function in relation to complex traits, as social behavior, is under current investigation (Hu and Barrett 2017). First evidence on the behavioral level indicates, for example, the potential role of epigenetics in the stress response system and associated behavioral differences such as risk-taking or novelty-seeking (Laviola et al. 2003; Kaminsky et al. 2008; Canestrelli et al. 2016). Also epigenetic mechanisms in the domain of memory formation and learning (Duke et al. 2017) may transfer to behavioral differences among individuals. But until we have replicated evidence of epigenetic effects on behavioral traits, a degree of humility about these findings would seem appropriate (see also Cobben and van Oers 2016). In particular, epigenetic explanations centering on specific genes should be interpreted carefully, as associations among single genes and behaviors often do not replicate in studies of humans and NHPs (see above). Given the increasing general understanding of genome-wide DNA methylation patterns in humans and NHPs (Lea et al. 2016, 2018), the role of epigenetics in personality development could become an interesting area of future research (Trillmich et al. 2018).

Summary and Outlook

We can infer that behavioral differences among individual NHPs develop towards increasing rank-order stability and a pattern indicative of what has been described as a "mature" personality in humans (but see exceptions in Manson and Perry 2013; Weiss and King 2015; Koski et al. 2017). Whereas environmental influences on behavioral variation among individuals act in humans especially around the time of

adolescence and young adulthood, behavioral variation in NHPs seems to already be affected early in life. Among these early environmental influences are stress-related variation in the natural environment, parenting style or rearing conditions. Later in life, migration or maternity during young adulthood may also affect personality development. As a complex interplay among genotype and the environment is likely, and the statistical power to detect even two-way interactions is low, current research is still far from disentangling the causal pathways that lead to behavioral differences. We propose that one possible way to peek inside this "black box" is to conduct genetically informed longitudinal studies or to use cross-sectional DOC modeling (Turkheimer and Harden 2014; Briley et al. 2018). That said, studies have to be adequately powered if they wish to use these tools. Since statistical power often turns out to be a problem in NHP studies, one possible direction might be to identify polygenic scores for behavioral differences in relatively large samples of a species, for example in breeding facilities, and then to apply this knowledge to the typically smaller populations in the wild or in other captive settings, such as zoos or sanctuaries. This could enable one to conduct genetically informative studies without the need for pedigree data or could supplement studies with (partly) existing pedigree data. Furthermore, testing evolutionary hypotheses stating under which conditions correlations among behavioral differences will occur and how stable these correlations are under changing environments or selection regimes (see Sih et al. 2004; Dochtermann and Dingemanse 2013) could be a fruitful direction for primate personality research. An example would be to test whether environmental variation affecting food resources favors different behavioral strategies or correlations among behaviors that form behavioral syndromes (Dingemanse et al. 2004). Human studies could also be informed, or inspired by, the increasing knowledge of dominance rank and hierarchy effects on behavioral variation in NHPs.

Compliance with Ethical Standards

Conflict of interest

We do not have any potential conflicts of interest to report.

References

Adams MJ (2014) Feasibility and uncertainty in behavior genetics for the nonhuman primates. Int J Primatol 35:156–168. https://doi.org/10.1007/s10764-013-9722-8

Adams MJ, King JE, Weiss A (2012) The majority of genetic variation in orangutan personality and subjective well-being is nonadditive. Behav Genet 42:675–686. https://doi.org/10.1007/s10519-012-9537-y

Anestis SF, Webster TH, Kamilar JM, Fontenot M B, Watts DP, Bradley BJ (2014) AVPR1A variation in chimpanzees (*Pan troglodytes*): Population differences and association with behavioral style. Int J Primatol 35:305–324. https://doi.org/10.1007/s10764-013-9747-z

Araya-Ajoy YG, Dingemanse NJ (2017) Repeatability, heritability, and age-dependence of seasonal plasticity in aggressiveness in a wild passerine bird. J Anim Ecol 86:227–238. https://doi.org/10.1111/1365-2656.12621

Arnaud CM, Suzumura T, Inoue E, Adams MJ, Weiss A, Inoue-Murayama M (2017) Genes, social transmission, but not maternal effects influence responses of wild Japanese macaques (*Macaca fuscata*) to novel-object and novel-food tests. Primates 58:103–113. https://doi.org/10.1007/s10329-016-0572-9

Bardi M, Bode AE, Ramirez SM, Brent LY (2005) Maternal care and development of stress responses in baboons. Am J Primatol 66:263–278. https://doi.org/10.1002/ajp.20143

Bardi M, Huffman MA (2002) Effects of maternal style on infant behavior in Japanese macaques (*Macaca fuscata*). Dev Psychobiol 41:364–372. https://doi.org/10.1002/dev.10065

Barr CS, Newman TK, Lindell S, Shannon C, Champoux M, Lesch KP, Suomi SJ, Goldman D, Higley JD (2004) Interaction between serotonin transporter gene variation and rearing condition in alcohol preference and consumption in female primates. Arch Gen Psychiat 61:1146-1152. https://doi.org/10.1001/archpsyc.61.11.1146

Page 116

Bell AM, Hankison SJ, Laskowski KL (2009) The repeatability of behaviour: a meta-analysis. Anim Behav 77: 771–783. https://doi.org/10.1016/j.anbehav.2008.12.022

Belsky J (2012) The development of human reproductive strategies: progress and prospects. Curr Dir Psychol Sci 21:310–316. https://doi.org/10.1177/0963721412453588

Berghänel A, Heistermann M, Schülke O, Ostner J (2016) Prenatal stress effects in a wild, long-lived primate: predictive adaptive responses in an unpredictable environment. Proc R Soc B 283:20161304. https://doi.org/10.1098/rspb.2016.1304

Berghänel A, Schulke O, Ostner J (2015) Locomotor play drives motor skill acquisition at the expense of growth: A life history trade-off. Sci Adv 1:e1500451. https://doi.org/10.1126/sciadv.1500451

Berman CM, Rasmussen KLR, Suomi SJ (1994) Responses of free-ranging rhesus monkeys to a natural form of social separation. I. Parallels with mother-infant separation in Ccaptivity. Child Dev 65:1028–1041. https://doi.org/10.1111/j.1467-8624.1994.tb00800.x

Bernstein IS (1976) Dominance, aggression and reproduction in primate societies. J Theor Biol 60:459–472. https://doi.org/10.1016/0022-5193(76)90072-2

Bierbach D, Laskowski KL, Wolf M (2017) Behavioural individuality in clonal fish arises despite near-identical rearing conditions. Nat Commun 8:15361. https://doi.org/10.1038/ncomms15361

Bleidorn W (2012) Hitting the road to adulthood: short-term personality development during a major life transition. Pers Soc Psychol B 38:1594–1608. https://doi.org/10.1177/0146167212456707

Bleidorn W, Kandler C, Caspi A (2014) The behavioural genetics of personality development in adulthood-classic, contemporary, and future trends. Eur J Personality 28:244–255. https://doi.org/10.1002/per.1957

Boake CRB (1989) Repeatability: Its role in evolutionary studies of mating behavior. Evol Ecol 3:173–182. https://doi.org/10.1007/BF02270919

Bolig R, Price CS, O'Neill PL, Suomi SJ (1992) Subjective assessment of reactivity level and personality traits of rhesus monkeys. Int J Primatol 13:287–306. https://doi.org/10.1007/BF02547817

Boyle EA, Li YI, Pritchard JK (2017) An expanded view of complex traits: From polygenic to omnigenic. Cell 169:1177–1186. https://doi.org/10.1016/j.cell.2017.05.038

Brent LJN, Melin AD (2014). The genetic basis of primate behavior: Genetics and genomics in field-based primatology. Int J Primatol 35:1–10. https://doi.org/10.1007/s10764-013-9732-6

Brent LJN, Semple S, MacLarnon A, Ruiz-Lambides A, Gonzalez-Martinez J, Platt ML (2014) Personality traits in rhesus macaques (*Macaca mulatta*) are heritable but do not predict reproductive output. Int J Primatol 35:188–209. https://doi.org/10.1007/s10764-013-9724-6

Briley DA, Livengood J, Derringer J (2018) Behaviour genetic frameworks of causal reasoning for personality psychology: behaviour genetics and causal reasoning. Eur J Personality 32:202–220. https://doi.org/10.1002/per.2153

Briley DA, Tucker-Drob EM (2017) Comparing the developmental genetics of cognition and personality over the life span: Comparing developmental genetics. J Pers 85:51–64. https://doi.org/10.1111/jopy.12186

Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, Skaug HJ, Machler M, Bolker BM (2017) Modeling zero-inflated count data with glmmTMB. bioRxiv, https://doi.org/10.1101/132753

Canestrelli D, Bisconti R, Carere C (2016) Bolder takes all? The behavioral dimension of biogeography. Trends Ecol Evol 31:35–43. https://doi.org/10.1016/j.tree.2015.11.004

Carter AJ, Feeney WE, Marshall HH, Cowlishaw G, Heinsohn R (2013) Animal personality: what are behavioural ecologists measuring? Biol Rev 88:465–475. https://doi.org/10.1111/brv.12007

Caspi A, McClay J, Moffitt TE, Mill J, Martin J, Craig IW, Taylor A, Poulton R (2002) Role of genotype in the cycle of violence in maltreated children. Science 297:851–854. https://doi.org/10.1126/science.1072290

Caspi A, Sugden K, Moffitt TE et al (2003) Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. Science 301:386–389. https://doi.org/10.1126/science.1083968

Chabris CF, Hebert BM, Benjamin DJ et al (2012) Most reported genetic associations with general intelligence are probably false positives. Psychol Sci 23:1314–1323. https://doi.org/10.1177/0956797611435528

Clarke AS, Boinski S (1995) Temperament in nonhuman primates. Am J Primatol 37:103–125. https://doi.org/10.1002/ajp.1350370205

Cobben MMP, van Oers K (2016) Bolder takes all and the role of epigenetics. A comment on Canestrelli et al. Trends Ecol Evol 31:498–499. https://doi.org/10.1016/j.tree.2016.04.007

De Lathouwers M, Van Elsacker L (2004) Comparing maternal styles in bonobos (*Pan paniscus*) and chimpanzees (*Pan troglodytes*). Am J Primatol 64:411–423. https://doi.org/10.1002/ajp.20089

Del Giudice M (2014) An evolutionary life history framework for psychopathology. Psychol Inq 25:261–300. https://doi.org/10.1080/1047840X.2014.884918

Denissen JJA, van Aken MAG, Penke L, Wood D (2013) Self-regulation underlies temperament and personality: an integrative developmental framework. Child Dev Perspect 7:255–260. https://doi.org/10.1111/cdep.12050

Digman JM (1990) Personality structure: emergence of the five-factor model. Annu Rev Psychol 41:417–440. https://doi.org/10.1146/annurev.ps.41.020190.002221

Dingemanse NJ, Both C, Drent PJ, Tinbergen JM (2004) Fitness consequences of avian personalities in a fluctuating environment. Proc R Soc Lond B 271:847–852. https://doi.org/10.1098/rspb.2004.2680

Dingemanse NJ, Dochtermann NA (2013) Quantifying individual variation in behaviour: mixed-effect modelling approaches. J Anim Ecol 82:39–54. https://doi.org/10.1111/1365-2656.12013

Dingemanse NJ, Kazem AJN, Réale D, Wright J (2010) Behavioural reaction norms: animal personality meets individual plasticity. Trends Ecol Evol 25:81–89. https://doi.org/10.1016/j.tree.2009.07.013

Dittus WPJ (1979) The evolution of behaviors regulating density and age-specific sex ratios in a primate population. Behaviour 69:265–301. https://doi.org/10.1163/156853979X00511

Dochtermann NA, Dingemanse NJ (2013) Behavioral syndromes as evolutionary constraints. Behav Ecol 24:806–811. https://doi.org/10.1093/beheco/art002

Donnellan MB, Conger RD, Burzette RG (2007). Personality development from late adolescence to young adulthood: Differential stability, normative maturity, and evidence for the maturity-stability hypothesis. J Pers 75:237–264. https://doi.org/10.1111/j.1467-6494.2007.00438.x

Duke CG, Kennedy AJ, Gavin CF, Day JJ, Sweatt JD (2017) Experience-dependent epigenomic

    reorganization in the hippocampus. Learn Memory 24:278–288.

    https://doi.org/10.1101/lm.045112.117

Fairbanks LA, McGuire MT (1993) Maternal protectiveness and response to the unfamiliar in vervet

    monkeys. Am J Primatol 30:119–129. https://doi.org/10.1002/ajp.1350300204

Fairbanks LA, Newman TK, Bailey JN, Jorgensen MJ, Breidenthal SE, Ophoff RA, Comuzzie AG, Martin LJ,

    Rogers J (2004) Genetic contributions to social impulsivity and aggressiveness in vervet monkeys.

    Biol Psychiat 55:642–647. https://doi.org/10.1016/j.biopsych.2003.12.005

Fawcett GL, Dettmer AM, Kay D, Raveendran M, Higley JD, Ryan ND, Cameron JL, Rogers J (2014)

    Quantitative genetics of response to novelty and other stimuli by infant rhesus macaques (*Macaca

    mulatta*) across three behavioral assessments. Int J Primatol 35:325–339.

    https://doi.org/10.1007/s10764-014-9750-z

Forss SIF, Koski SE, van Schaik CP (2017) Explaining the paradox of neophobic explorers: the social

    information hypothesis. Int J Primatol 38:799–822. https://doi.org/10.1007/s10764-017-9984-7

Freeman HD, Weiss A, Ross SR (2016) Atypical early histories predict lower extraversion in captive

    chimpanzees: early histories predict extraversion. Dev Psychobiol 58:519–527.

    https://doi.org/10.1002/dev.21395

French JA (1981) Individual differences in play in *Macaca fuscata*: The role of maternal status and proximity.

    Int J Primatol 2:237–246. https://doi.org/10.1007/BF02739332

Gillespie NA, Zhu G, Neale MC, Heath AC, Martin NG (2003) Direction of causation modeling between

    cross-sectional measures of parenting and psychological distress in female twins. Behav Genet

    33:383–396

Gosling SD (2001) From mice to men: what can we learn about personality from animal research? Psychol

    Bull 127:45–86

Gutleb DR, Roos C, Noll A, Ostner J, Schülke O (2018) COMT Val[158] Met moderates the link between

    rank and aggression in a non-human primate. Genes Brain Behav 17:e12443.

    https://doi.org/10.1111/gbb.12443

Harlow HF, Dodsworth RO, Harlow MK (1965) Total social isolation in monkeys. P Natl Acad Sci USA 54:90–97

Hill WD, Arslan RC, Xia C et al (2018) Genomic analysis of family data reveals additional genetic effects on intelligence and personality. Mol Psychiatr 23:2347-2362. https://doi.org/10.1038/s41380-017-0005-1

Hopkins WD, Keebaugh AC, Reamer LA, Schaeffer J, Schapiro SJ, Young LJ (2015) Genetic influences on receptive joint attention in chimpanzees (*Pan troglodytes*). Sci Rep 4:3774. https://doi.org/10.1038/srep03774

Hopkins WD, Reamer L, Mareno MC, Schapiro SJ (2014) Genetic basis in motor skill and hand preference for tool use in chimpanzees (*Pan troglodytes*). Proc R Soc B 282:20141223. https://doi.org/10.1098/rspb.2014.1223

Hu J, Barrett RDH (2017) Epigenetics in natural animal populations. J Evol Biol 30:1612–1632. https://doi.org/10.1111/jeb.13130

Hutteman R, Nestler S, Wagner J, Egloff B, Back MD (2015) Wherever I may roam: Processes of self-esteem development from adolescence to emerging adulthood in the context of international student exchange. J Pers Soc Psychol 108:767–783. https://doi.org/10.1037/pspp0000015

Inoue-Murayama M, Yokoyama C, Yamanashi Y, Weiss A (2018) Common marmoset (*Callithrix jacchus*) personality, subjective well-being, hair cortisol level and *AVPR1a*, *OPRM1*, and *DAT* genotypes. Sci Rep 8:10255. https://doi.org/10.1038/s41598-018-28112-7

Johnson W (2014) Developing Difference. Palgrave Macmillan, Basingstoke

Johnson W, Deary IJ, Iacono WG (2009) Genetic and environmental transactions underlying educational attainment. Intelligence 37:466–478. https://doi.org/10.1016/j.intell.2009.05.006

Johnson W, Penke L, Spinath FM (2011) Heritability in the era of molecular genetics: Some thoughts for understanding genetic influences on behavioural traits. Eur J Personality 25:254–266. https://doi.org/10.1002/per.836

Johnson Z, Brent L, Alvarenga JC et al (2015) Genetic influences on response to novel objects and dimensions of personality in *Papio* baboons. Behav Genet 45:215–227. https://doi.org/10.1007/s10519-014-9702-6

Kain JS, Stokes C, de Bivort BL (2012) Phototactic personality in fruit flies and its suppression by serotonin and white. P Natl Acad Sci USA 109:19834–19839. https://doi.org/10.1073/pnas.1211988109

Kaminsky Z, Petronis A, Wang S-C, Levine B, Ghaffar O, Floden D, Feinstein A (2008) Epigenetics of personality traits: an illustrative study of identical twins discordant for risk-taking behavior. Twin Res Hum Genet 11:1–11. https://doi.org/10.1375/twin.11.1.1

Kandler C, Bleidorn W, Riemann R, Angleitner A, Spinath FM (2012) Life events as environmental states and genetic traits and the role of personality: A longitudinal twin study. Behav Genet 42:57–72. https://doi.org/10.1007/s10519-011-9491-0

King J E, Figueredo AJ (1997) The five-factor model plus dominance in chimpanzee personality. J Res Pers 31:257–271. https://doi.org/10.1006/jrpe.1997.2179

King JE, Weiss A, Sisco MM (2008) Aping humans: Age and sex effects in chimpanzee (*Pan troglodytes*) and human (*Homo sapiens*) personality. J Comp Psychol 122:418–427. https://doi.org/10.1037/a0013125

Kohn JN, Snyder-Mackler N, Barreiro LB, Johnson ZP, Tung J, Wilson ME (2016) Dominance rank causally affects personality and glucocorticoid regulation in female rhesus macaques. Psychoneuroendocrinol 74:179–188. https://doi.org/10.1016/j.psyneuen.2016.09.005

Koski SE (2011) How to measure animal personality and why does it matter? Integrating the psychological and biological approaches to animal personality. In: Inoue-Murayama M, Kawamura S, Weiss A (eds) From Genes to Animal Behavior. Springer Japan, Tokyo, pp 115-136

Koski SE, Buchanan-Smith HM, Ash H, Burkart JM, Bugnyar T, Weiss A (2017) Common marmoset (*Callithrix jacchus*) personality. J Comp Psychol 131:326–336. https://doi.org/10.1037/com0000089

Koyama N (1970) Changes in dominance rank and division of a wild Japanese monkey troop in Arashiyama. Primates 11:335–390. https://doi.org/10.1007/BF01730638

Krawczak M, Trefilov A, Berard J, Bercovitch F, Kessler M, Sauermann U, Croucher P, Nürnberg P, Widdig A, Schmidtke J (2005) Male reproductive timing in rhesus macaques is influenced by the 5HTTLPR promoter polymorphism of the serotonin transporter gene 1. Biol Reprod 72:1109–1113. https://doi.org/10.1095/biolreprod.104.038059

Krueger RF, South S, Johnson W, Iacono W (2008) The heritability of personality is not always 50%: Gene-environment interactions and correlations between personality and parenting. J Pers 76:1485–1522. https://doi.org/10.1111/j.1467-6494.2008.00529.x

Kulik L, Amici F, Langos D, Widdig A (2015a) Sex differences in the development of aggressive behavior in rhesus macaques (*Macaca mulatta*). Int J Primatol 36:764–789. https://doi.org/10.1007/s10764-015-9853-1

Kulik L, Amici F, Langos D, Widdig A (2015b) Sex differences in the development of social relationships in rhesus macaques (*Macaca mulatta*). Int J Primatol 36:353–376. https://doi.org/10.1007/s10764-015-9826-4

Latzman RD, Freeman HD, Schapiro SJ, Hopkins WD (2015) The contribution of genetics and early rearing experiences to hierarchical personality dimensions in chimpanzees (*Pan troglodytes*). J Pers Soc Psychol 109:889–900. https://doi.org/10.1037/pspp0000040

Laviola G, Macrì S, Morley-Fletcher S, Adriani W (2003) Risk-taking behavior in adolescent mice: psychobiological determinants and early epigenetic influence. Neurosci Biobehav R 27:19–31. https://doi.org/10.1016/S0149-7634(03)00006-X

Lea AJ, Altmann J, Alberts SC, Tung J (2015) Developmental constraints in a wild primate. Am Nat 185:809–821. https://doi.org/10.1086/681016

Lea AJ, Altmann J, Alberts SC, Tung J (2016) Resource base influences genome-wide DNA methylation levels in wild baboons (*Papio cynocephalus*). Mol Ecol 25:1681–1696. https://doi.org/10.1111/mec.13436

Lea AJ, Vockley CM, Johnston RA, Del Carpio CA, Barreiro LB, Reddy TE, Tung J (2017) Genome-wide quantification of the effects of DNA methylation on human gene regulation. bioRxiv, https://doi.org/10.1101/146829

Madlon-Kay S, Montague MJ, Brent LJN, Ellis S, Zhong B, Snyder-Mackler N, Horvath JE, Skene JHP, Platt ML (2018) Weak effects of common genetic variation in oxytocin and vasopressin receptor genes on rhesus macaque social behavior. Am J Primatol 80:e22873. https://doi.org/10.1002/ajp.22873

Madrid JE, Mandalaywala TM, Coyne SP, Ahloy-Dallaire J, Garner JP, Barr CS, Maestripieri D, Parker KJ (2018) Adaptive developmental plasticity in rhesus macaques: the serotonin transporter gene interacts with maternal care to affect juvenile social behaviour. Proc R Soc B 285:20180541

Maestripieri D (2018) Maternal influences on primate social development. Behav Ecol Sociobiol 72:130. https://doi.org/10.1007/s00265-018-2547-x

Maestripieri D, McCormack K, Lindell SG, Higley JD, Sanchez MM (2006) Influence of parenting style on the offspring's behaviour and CSF monoamine metabolite levels in crossfostered and noncrossfostered female rhesus macaques. Behav Brain Res 175:90–95. https://doi.org/10.1016/j.bbr.2006.08.002

Manson JH, Perry S (2013) Personality structure, sex differences, and temporal change and stability in wild white-faced capuchins (*Cebus capucinus*). J Comp Psychol 127:299–311. https://doi.org/10.1037/a0031316

Martin JE (2005) The influence of rearing on personality ratings of captive chimpanzees (*Pan troglodytes*). Appl Anim Behav Sci 90:167–181. https://doi.org/10.1016/j.applanim.2004.08.019

Martin JS, Suarez SA (2017) Personality assessment and model comparison with behavioral data: A statistical framework and empirical demonstration with bonobos (*Pan paniscus*). Am J Primatol 79:e22670. https://doi.org/10.1002/ajp.22670

Massen JJM, Antonides A, Arnold A-M K, Bionda T, Koski SE (2013) A behavioral view on chimpanzee personality: Exploration tendency, persistence, boldness, and tool-orientation measured with group experiments. Am J Primatol 75:947–958. https://doi.org/10.1002/ajp.22159

Mendle J, Harden K P, Turkheimer E, Van Hulle CA, D'Onofrio BM, Brooks-Gun J, Rodgers JL, Emery RE, Lahey BB (2009) Associations between father absence and age of first sexual intercourse. Child Dev 80:1463–1480. https://doi.org/10.1111/j.1467-8624.2009.01345.x

Moffitt TE (2005) The new look of behavioral genetics in developmental psychopathology: Gene-environment interplay in antisocial behaviors. Psychol Bull 131:533–554. https://doi.org/10.1037/0033-2909.131.4.533

Montiglio P-O, Ferrari C, Réale D (2013) Social niche specialization under constraints: personality, social

    interactions and environmental heterogeneity. Phil Trans R Soc B 368:20120343.

    https://doi.org/10.1098/rstb.2012.0343

Munafò MR, Flint J (2004) Meta-analysis of genetic association studies. Trends Genet 20:439–444.

    https://doi.org/10.1016/j.tig.2004.06.014

Munafò MR, Flint J (2011) Dissecting the genetic architecture of human personality. Trends Cogn Sci

    15:395-400. https://doi.org/10.1016/j.tics.2011.07.007

Munafò MR, Zammit S, Flint J (2014) Practitioner review: a critical perspective on gene-environment

    interaction models - what impact should they have on clinical perceptions and practice? J Child

    Psychol Psychiat 55:1092–1101. https://doi.org/10.1111/jcpp.12261

Murray CM, Stanton MA, Wellens KR, Santymire RM, Heintz MR, Lonsdorf EV (2018) Maternal effects on

    offspring stress physiology in wild chimpanzees: chimpanzee maternal effects on offspring stress. Am J

    Primatol 80:e22525. https://doi.org/10.1002/ajp.22525

Nakagawa S, Schielzeth H (2010) Repeatability for Gaussian and non-Gaussian data: a practical guide for

    biologists. Biol Rev 85:935-956. https://doi.org/10.1111/j.1469-185X.2010.00141.x

Neumann C, Agil M, Widdig A, Engelhardt A (2013) Personality of wild male crested macaques (*Macaca*

    nigra). PLoS ONE 8:e69383. https://doi.org/10.1371/journal.pone.0069383

Newman TK, Syagailo YV, Barr CS, Wendland JR, Champoux M, Graessle M, Suomi SJ, Higley JD, Lesch K-

    P (2005) Monoamine oxidase A gene promoter variation and rearing experience influences aggressive

    behavior in rhesus monkeys. Biol Psychiat 57:167–172. https://doi.org/10.1016/j.biopsych.2004.10.012

Neyer F J, Lehnart J (2007) Relationships matter in personality development: evidence from an 8-year

    longitudinal study across young adulthood. J Pers 75:535–568. https://doi.org/10.1111/j.1467-

    6494.2007.00448.x

Okbay A, Beauchamp, JP, Fontana MA et al (2016) Genome-wide association study identifies 74 loci

    associated with educational attainment. Nature 533:539–542. https://doi.org/10.1038/nature17671

Plomin R, Bergeman CS (1991) The nature of nurture: Genetic influence on "environmental" measures.

    Behav Brain Sci 14:373–386. https://doi.org/10.1017/S0140525X00070278

Plomin R, DeFries JC, Knopik VS, Neiderhiser JM (2012) Behavioral Genetics, 6th edn. Worth Publishers, New York, NY

Plomin R, von Stumm S (2018) The new genetics of intelligence. Nat Rev Genet 19:148–159. https://doi.org/10.1038/nrg.2017.104

Réale D, Reader SM, Sol D, McDougall PT, Dingemanse NJ (2007) Integrating animal temperament within ecology and evolution. Biol Rev 82:291–318. https://doi.org/10.1111/j.1469-185X.2007.00010.x

Roberts BW, DelVecchio WF (2000) The rank-order consistency of personality traits from childhood to old age: a quantitative review of longitudinal studies. Psych Bull 126:3–25

Roberts BW, Helson R, Klohnen EC (2002) Personality development and growth in women across 30 years: three perspectives. J Pers 70:79–102

Roberts BW, Walton KE, Viechtbauer W (2006) Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. Psych Bull 132:1–25. https://doi.org/10.1037/0033-2909.132.1.1

Roche DG, Careau V, Binning SA (2016) Demystifying animal "personality" (or not): why individual variation matters to experimental biologists. J Exp Biol 219:3832–3843. https://doi.org/10.1242/jeb.146712

Rogers J, Shelton SE, Shelledy W, Garcia R, Kalin NH (2008) Genetic influences on behavioral inhibition and anxiety in juvenile rhesus macaques. Genes Brain Behav 7:463–469. https://doi.org/10.1111/j.1601-183X.2007.00381.x

Runcie DE, Wiedmann RT, Archie EA, Altmann J, Wray GA, Alberts SC, Tung J (2013) Social environment influences the relationship between genotype and gene expression in wild baboons. Phil Trans R Soc B 368:20120345. https://doi.org/10.1098/rstb.2012.0345

Rushton JP, Brainerd CJ, Pressley M (1983) Behavioral development and construct validity: The principle of aggregation. Psych Bull 94:18–38. https://doi.org/10.1037/0033-2909.94.1.18

Sallis H, Davey Smith G, Munafò MR (2018) Genetics of biologically based psychological differences. Phil Trans R Soc B 373:20170162. https://doi.org/10.1098/rstb.2017.0162

Sanchez-Roige S, Gray JC, MacKillop J, Chen C-H, Palmer AA (2018) The genetics of human personality. Genes Brain Behav 17:12439. https://doi.org/10.1111/gbb.12439

Sapolsky RM, Share LJ (2004) A pacific culture among wild baboons: Its emergence and transmission. PLoS Biol 2:e106. https://doi.org/10.1371/journal.pbio.0020106

Scarr S, McCartney K (1983) How people make their own environments: a theory of genotype greater than environment effects. Child Dev 54:424–435

Sih A, Bell A, Johnson JC (2004) Behavioral syndromes: an ecological and evolutionary overview. Trends Ecol Evol 19:372–378. https://doi.org/10.1016/j.tree.2004.04.009

Sih A, Mathot KJ, Moirón M, Montiglio P-O, Wolf M, Dingemanse NJ (2015) Animal personality and state–behaviour feedbacks: a review and guide for empiricists. Trends Ecol Evol 30:50–60. https://doi.org/10.1016/j.tree.2014.11.004

Silk JB (2002) Kin selection in primate groups. Int J Primatol 23:849-875

Staes N, Koski SE, Helsen P, Fransen E, Eens M, Stevens JMG (2015) Chimpanzee sociability is associated with vasopressin (Avpr1a) but not oxytocin receptor gene (OXTR) variation. Horm Behav 75:84–90. https://doi.org/10.1016/j.yhbeh.2015.08.006

Staes N, Weiss A, Helsen P, Korody M, Eens M, Stevens JMG (2016) Bonobo personality traits are heritable and associated with vasopressin receptor gene 1a variation. Sci Rep 6:38193. https://doi.org/10.1038/srep38193

Stamps J, Groothuis TGG (2010) The development of animal personality: relevance, concepts and perspectives. Biol Rev 85:301–325. https://doi.org/10.1111/j.1469-185X.2009.00103.x

Stevenson-Hinde J, Stillwell-Barnes R, Zunz M (1980) Individual differences in young rhesus monkeys: Consistency and change. Primates 21:498–509. https://doi.org/10.1007/BF02373838

Stevenson-Hinde J, Zunz M (1978) Subjective assessment of individual rhesus monkeys. Primates 19:473–482. https://doi.org/10.1007/BF02373309

Suomi SJ (1997) Early determinants of behaviour: evidence from primate studies. Brit Med Bull 53:170–184

Suomi SJ (2004) How gene-environment interactions shape biobehavioral development: Lessons from studies with rhesus monkeys. Res Hum Dev 1:205–222. https://doi.org/10.1207/s15427617rhd0103_5

Sussman AF, Mates EA, Ha JC, Bentson KL, Crockett CM (2014) Tenure in current captive setting and age predict personality changes in adult pigtailed macaques. Anim Behav 89:23–30. https://doi.org/10.1016/j.anbehav.2013.12.009

Terracciano A, Costa PT, McCrae RR (2006) Personality plasticity after age 30. Pers Soc Psychol Bull 32:999–1009. https://doi.org/10.1177/0146167206288599

Trefilov A, Berard J, Krawczak M, Schmidtke J (2000) Natal dispersal in rhesus macaques is related to serotonin transporter gene promoter variation. Behav Genet 30:295–301

Trillmich F, Müller T, Müller C (2018) Understanding the evolution of personality requires the study of mechanisms behind the development and life history of personality traits. Biol Lett 14:20170740. https://doi.org/10.1098/rsbl.2017.0740

Tucker-Drob EM, Briley DA (2019) Theoretical concepts in the genetics of personality development. In: McAdams DP, Shiner RL, Tackett JL (eds) The handbook of personality development. Guilford, New York, pp 40-58.

Turkheimer E (2000) Three laws of behavior genetics and what they mean. Curr Dir Psychol Sci 9:160–164. https://doi.org/10.1111/1467-8721.00084

Turkheimer E, Harden KP (2013) Behavior genetic research methods. In: Reis HT, Judd CM (eds) Handbook of Research Methods in Social and Personality Psychology, 2nd edn. Cambridge University Press, New York, pp 159–187

Turkheimer E, Pettersson E, Horn EE (2014) A phenotypic null hypothesis for the genetics of personality. Annu Rev Psychol 65:515–540. https://doi.org/10.1146/annurev-psych-113011-143752

Uher J (2008) Comparative personality research: methodological approaches. Eur J Personality 22:427–455. https://doi.org/10.1002/per.680

Uher J (2011) Individual behavioral phenotypes: An integrative meta-theoretical framework. Why "behavioral syndromes" are not analogs of "personality." Dev Psychobiol 53:521–548. https://doi.org/10.1002/dev.20544

Uher J, Addessi E, Visalberghi E (2013) Contextualised behavioural measurements of personality differences obtained in behavioural tests and social observations in adult capuchin monkeys (*Cebus apella*). J Res Pers 47:427–444. https://doi.org/10.1016/j.jrp.2013.01.013

van den Berg SM, de Moor MHM, McGue M et al (2014) Harmonization of neuroticism and extraversion phenotypes across inventories and cohorts in the genetics of personality consortium: an application of item response theory. Behav Genet 44:295–313. https://doi.org/10.1007/s10519-014-9654-x

van IJzendoorn MH, Bard KA, Bakermans-Kranenburg MJ, Ivan K (2009) Enhancement of attachment and cognitive development of young nursery-reared chimpanzees in responsive versus standard care. Dev Psychobiol 51:173–185. https://doi.org/10.1002/dev.20356

von Borell C, Kulik L, Widdig A (2016) Growing into the self: the development of personality in rhesus macaques. Anim Behav 122:183–195. https://doi.org/10.1016/j.anbehav.2016.10.013

Vukasović T, Bratko D (2015) Heritability of personality: A meta-analysis of behavior genetic studies. Psychol Bull 141:769–785. https://doi.org/10.1037/bul0000017

Watson KK, Li D, Brent LJN, Horvath JE, Gonzalez-Martinez J, Ruíz-Lambides AV, Robinson AG, Skene JH, Platt ML (2015) Genetic influences on social attention in free-ranging rhesus macaques. Anim Behav 103:267–275. https://doi.org/10.1016/j.anbehav.2015.02.012

Weiss A (2017a) A human model for primate personality. Proc R Soc B 284:20171129. https://doi.org/10.1098/rspb.2017.1129

Weiss A (2017b) Exploring factor space (and other adventures) with the hominoid personality questionnaire. In: Vonk J, Weiss A, Kuczaj SA (eds) Personality in Nonhuman Animals. Springer International Publishing, Cham, pp 19–38

Weiss A, Adams MJ, Widdig A, Gerald MS (2011) Rhesus macaques (*Macaca mulatta*) as living fossils of hominoid personality and subjective well-being. J Comp Psychol 125:72–83. https://doi.org/10.1037/a0021187

Weiss A, Inoue-Murayama M, Hong K-W, Inoue E, Udono T, Ochiai T, Matsuzawa T, Hirata S, King JE (2009) Assessing chimpanzee personality and subjective well-being in Japan. Am J Primatol 71:283–292. https://doi.org/10.1002/ajp.20649

Weiss A, King JE (2015) Great ape origins of personality maturation and sex differences: A study of
    orangutans and chimpanzees. J Pers Soc Psychol 108:648–664. https://doi.org/10.1037/pspp0000022

Weiss A, King JE, Figueredo AJ (2000) The heritability of personality factors in chimpanzees (*Pan
    troglodytes*). Behav Genet 30:213–221

Weiss A, Staes N, Pereboom JJM, Inoue-Murayama M, Stevens JMG, Eens M (2015) Personality in bonobos.
    Psychol Sci 26:1430–1439. https://doi.org/10.1177/0956797615589933

Weiß BM, Kulik L, Ruiz-Lambides AV, Widdig A (2016) Individual dispersal decisions affect fitness via
    maternal rank effects in male rhesus macaques. Sci Rep 6:32212. https://doi.org/10.1038/srep32212

Wilson VAD, Weiss A, Humle T, Morimura N, Udono T, Idani G, Matsuzawa T, Hirata S, Inoue-Murayama
    M (2017) Chimpanzee personality and the arginine vasopressin receptor 1A genotype. Behav Genet
    47:215–226. https://doi.org/10.1007/s10519-016-9822-2

Wolf M, van Doorn GS, Weissing FJ (2008) Evolutionary emergence of responsive and unresponsive
    personalities. P Natl Acad Sci USA 105:15825–15830. https://doi.org/10.1073/pnas.0805473105

Yamamoto ME, Lopes FA (2004) Effect of removal from the family group on feeding behavior by captive
    Callithrix jacchus. Int J Primatol 25:489–500. https://doi.org/10.1023/B:IJOP.0000019164.98756.9c

Zimmermann J, Neyer FJ (2013) Do we become a different person when hitting the road? Personality
    development of sojourners. J Pers Soc Psychol 105:515–530. https://doi.org/10.1037/a0033019

Zuur AF, Ieno EN, Walker N, Saveliev AA, Smith GM (2009) Mixed effects models and extensions in
   ecology with R. Springer, New York, NY

Figure Captions

Fig. 1 The difference between mean-level and rank-order stability

Scenario a): The rank-order of differences in aggressiveness stays stable between all four individuals
throughout development, while the mean-level aggressiveness in the population increases with age.
Scenario b): The mean-level aggressiveness in the population stays stable throughout development, while
the rank-order of aggressiveness changes between the four individuals over the years.

Figure 1

a)

b)
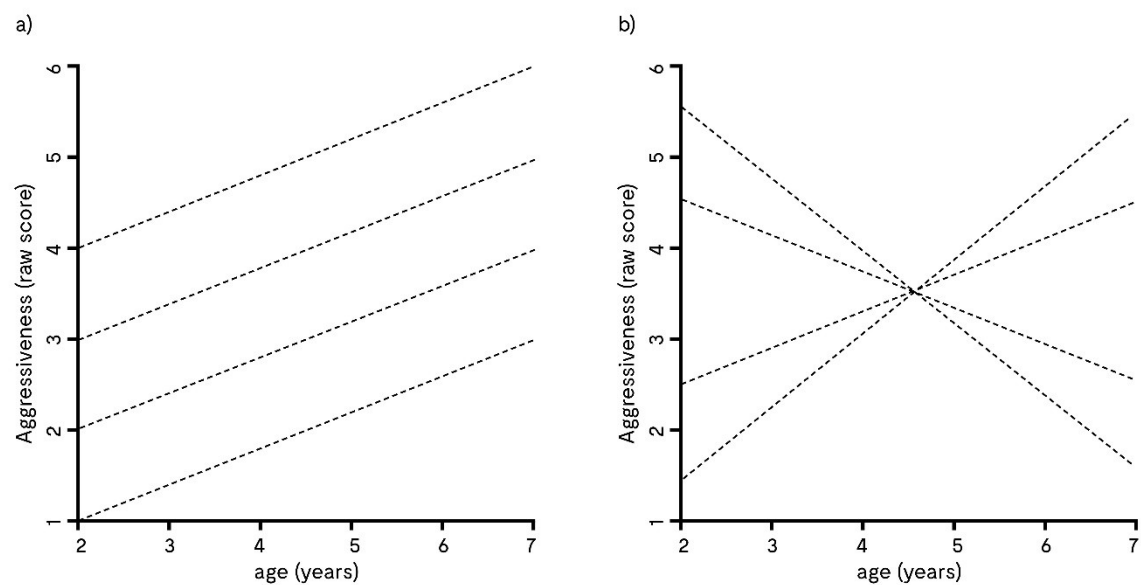
Appendix C: Manuscript III

An integrative study of facultative personality calibration

The theory of facultative calibration, which explains personality differences as responses to variation in other phenotypic traits of individuals, received mixed results throughout the last years. Whereas there is strong evidence that individual differences in human behavior are correlated with the self-perception of other traits, it still needs to be questioned whether they are also adjusted to objective differences in body condition (i.e. formidability). In two independent studies (N_1 = 119 men and 124 women, N_2 = 165 men) we tested hypotheses of facultative personality calibration in an integrative way, assessing various outcomes of previous studies in the same samples (including Anger Proneness, Extraversion, Neuroticism, Narcissism, Shyness, Vengefulness, and Sociosexual Orientation). Formidability was derived from assessments of physical strength and various anthropometric measures from full-body 3D scans and paired with measures of self-perceived and other-rated physical attractiveness (based on rotating morphometric 3D body models and facial photographs). We could replicate positive correlations with self-perceived attractiveness across outcomes, though these were not corroborated by more objective assessments of attractiveness: an effect of other-rated attractiveness was clearly not supported in our results for either sex, regardless of the personality outcome. Anthropometric measures and physical strength were also largely unrelated to personality, with the exception of Extraversion, Utility of Personal Aggression, and Sociosexual Orientation. While the two samples differed in their results for domain-level Extraversion, at least the Extraversion facets Activity and Assertiveness were related to strength and masculinity in men. For Sociosexual Orientation the results of our two samples varied more substantially, a positive association was only present in Study 2. Future studies need to clarify whether formidability, potentially an indicator of genetic quality for males, enhances their orientation and success in short-term mating. Furthermore we propose longitudinal twin-difference studies as means to evaluate the theory of personality recalibration in a more controlled manner.

Keywords: facultative calibration, personality traits, reactive heritability, relative bargaining power, sociosexual orientation, state-behavior feedback, anthropometrics, physical attractiveness

1. Introduction

Throughout the last decades different theories emerged that explain personality variation among individuals from an evolutionary perspective (Penke et al., 2007; Buss & Hawley, 2010; Buss & Penke, 2015). One of them, the theory of facultative calibration, poses that individual differences in personality may not be actively selected for, but are calibrated to differences in other traits of individuals, such as physical strength, physical attractiveness, or intelligence (Haysom et al., 2015; Lukaszewski & Roney, 2011; Sell et al., 2009; Tooby & Cosmides, 1990). These traits are thought to enhance the formidability or expected relative bargaining power (RBP) of individuals, i.e. the ability to inflict costs on others or to extract benefits from them (Petersen et al., 2010). The heritability of personality should then be of reactive manner, linked to the heritability of these specific traits. Notably, such explanation of personality variation is not exclusive to humans. In animal behavioral ecology, a similar idea has been put forward as "state-behavior feedback loop", reflecting that behavior and behavioral repeatability (the stability of a behavioral trait throughout time) may be adaptively adjusted to slower-changing or fixed state variables such as size, energy reserves, or parasite infection (Sih et al., 2015; Wolf & Weissing, 2010). During the last years, various studies of humans and non-human animals empirically tested such links between behavioral and other phenotypic traits, however not necessarily under the same label. When we subsequently refer to the theory of facultative calibration in humans, we intend it to subsume synonymous terms as "recalibrational theory" or "condition-dependent calibration" that have been used in the literature before.

One of the first studies that explicitly tested facultative calibration in humans predicted a relationship between RBP and anger proneness (Sell et al., 2009). It was reasoned that differences in RBP would lead to differences in the perception of a personal welfare trade-off ratio WTR (i.e., how much an actor A would expect an actor B to value his welfare compared to B's own; Tooby et al., 2008), which in turn should affect the propensity to feel and express anger (see Sell et al., 2009). Furthermore, the authors predicted sex differences in which variables would factor into an individual`s RBP. Since men are usually stronger and tend to monopolize the use of force in social negotiations, strength should factor strongly into men's RBP (by the means of increased physical formidability; see Lukaszewski, 2013). In women, RBP should be more closely linked to physical attractiveness, a main indicator of fertility and reproductive potential.

Given that access of males to female sexuality is more limited than the other way round, attractiveness should thus provide a powerful leverage to women. Indeed, Sell et al. (2009) found upper body strength to be positively correlated with a variety of anger-relevant measures in men, but not in women. Self-perceived attractiveness, in turn, showed positive relationships to anger in females, but only a few such associations were found in men.

Price et al. (2012) extended these results by linking anger proneness to a greater variety of anthropometric measures (e.g. chest circumference, bicep circumference, waist circumference, and overall body shape masculinity). They found measures of body shape and upper body masculinity in particular to be positively related to proneness to anger in men. However, this pattern was merely evident in a subsample of younger participants aged 18 to 23. In contrast, neither physical masculinity, nor anthropometric measures of attractiveness (such as waist-hip ratio or BMI) were related to proneness to anger in female participants, with the exception of leg-body ratio (LBR). Nevertheless they could replicate the results of Sell et al. (2009) showing a positive relationship between different measures of self-perceived attractiveness and anger for females (but not for males). Recent findings from a large sample of Swiss adolescents ($N = 1447$; Sell, Eisner, & Ribeaud, 2016) found no correlation between height and aggressiveness, but a small relationship of aggressiveness with weight and BMI. However, in multiple regressions also including overall fighting ability (a composite measure based on self-report items and flexed biceps circumference), these effects disappeared. Also, a single effect of biceps circumference was no longer significant when controlling for the self-report measure of fighting ability. Hence the authors conclude that actually fighting ability predicts aggressive bargaining rather than individual anthropometric measures per se. Archer and Thanzami (2007) suggested a more fine-grained differentiation in anger and aggression related outcomes. While they found a relationship between trait measures of direct (physical) aggressiveness with height, weight, and strength in a non-western sample of young Indian men, they did not find a correlation between physical formidability and proneness to anger. In fact, based on the concept of Resource Holding Power (i.e., the ability to win a fight; Parker, 1974; Stulp et al., 2012) they explicitly hypothesized size and strength to be unrelated to indirect measures of aggressiveness such as anger or hostility. Furthermore, it has been argued that vengefulness, the propensity to harm others or withhold benefits *in response* to a previous cost-

inflicting or benefit withholding event, could be adjusted based on mechanism of facultative calibration as well (McCullough et al., 2011).

Lukazsewski and Roney (2011) hypothesized facultative calibration of extraversion, arguing that since extraverted individuals, as compared to more introverted individuals, are more likely to proactively seek social status, influence, and relationships, they will be exposed more often to conflicts of interest with others. The entailed cost-benefit ratio should thus be more favorable to stronger individuals and also to more attractive individuals, since they are known to be preferred in relationships and cooperative exchanges. However, they did not predict an isolated effect of facultative calibration but an integrative model wherein facultative calibration and a pleiotropic genetic effect acting on both extraversion and somatic features influence individual variability in extraversion together. In fact, they found extraversion to be related to other-rated- and self-perceived attractiveness in both sexes, to physical strength in men and independently to a polymorphism in the androgen receptor gene in men (which had been previously linked to strength and extraversion, though overall results are mixed). Their results therefore imply that the heritability of extraverted personality is comprised of both the influence of genetic polymorphisms acting (rather) directly on behavioral trait regulation and of reactive heritability reflected in facultative calibration. However, in a subsequent study, Lukaszewski (2013) could not replicate a relationship between other-rated attractiveness and various personality traits (e.g. extraversion, emotionality, and fear of rejection) in either sex. Physical strength was again related to extraversion, this time in both sexes. Similar results have been found in the Tsimane, a group of forager-horticulturalists living in the Bolivian amazon (von Rueden et al., 2015). In this study, physical strength in both sexes was related to the personality dimension of Prosocial Leadership Orientation, which is specific to the Tsimane and represents a mixture of high Extraversion, high Agreeableness, high Openness to Experience, and low Neuroticism (based on items of the Big Five Inventory). Physical strength explained about 15% of the additive heritability of Prosocial Leadership Orientation (von Rueden et al., 2015).

Furthermore, narcissism, a complex psychological trait typically correlated to extraversion which includes feelings of superiority, entitlement, and power (Paulhus & Williams, 2002; Wetzel et al., 2016), has been shown to form an instantiation of a personality-formidability correlation, as it is associated with

physical attractiveness (Holtzman & Strube, 2010). On the one hand, a correlation between narcissism and physical attractiveness could be explained in evolutionary terms by a convergence of selection pressures on narcissistic traits and physical attractiveness in short-term mating contexts (Holtzman & Strube, 2010). On the other hand, the proposed association also dovetails with the notion of a relationship between anger proneness and formidability / RBP, since the sense of entitlement is (a) involved in setting a WTR, influencing subsequent expressions of anger (Sell et al., 2009), and (b) constitutes a core component of the narcissistic personality disposition (Ackerman et al., 2011; Campbell et al., 2010). In a meta-analysis, Holtzman and Strube (2010) calculated a mean correlation of 0.14 between measures of narcissism and other-rated physical attractiveness.

Finally, facultative calibration has been proposed as a mechanism of individual differences in sociosexual orientation. Lukaszewski et al. (2014) found positive correlations of composite scores of physical strength and attractiveness (comprising both self- and other-rated measures) with an uncommitted (short-term) mating orientation in men. These relationships were predicted to emerge because strength and attractiveness were assumed to serve as cues of genetic quality and were thus likely to be preferred by ancestral women in uncommitted mating. Additionally, physical strength was most likely an asset in intrasexual contests arising in this setting. Such relationships were neither predicted nor found for women (Lukaszewski et al., 2014).

Notably, several studies have challenged the theory of facultative personality calibration in recent years. For instance, a longitudinal study on the ontogeny of aggressiveness in children showed that boys with greater aggressive and antisocial tendencies at age 11 had greater increases in physical strength during the following six years of puberty, while not being consistently stronger than their peers at age 11 (Isen et al., 2015). These findings are not consistent with the theory of facultative calibration. Thus, Isen et al. (2015) proposed a joint hormonal mediation of behavioral and physiological traits as an alternative explanation. Haysom et al. (2015) found no correlation between extraversion and height or BMI in men or women of a large twin sample (N=1659). In addition, low but significant phenotypic correlations between extraversion and facial attractiveness were not genetically mediated in this study and could also be explained by general learning processes (Haysom et al., 2015). Overall, the theoretical foundations of facultative calibration have

been discussed in a critical light by Zietsch (2016), since for example the large mutational target size of complex behavioral traits likely affects personality variation, or strategies counter to previously proposed mechanisms could be at work (e.g. physically unattractive men may try to attract females with extraverted behavior). However, in some points, fairness must be maintained with previous studies who empirically tested possible explanations of the proposed optimal strategy (e.g. by showing a correlation between self-perceived bargaining power and fear of rejection; Lukaszewski, 2013) or did simply not claim that facultative calibration is the only mechanism driving variation in personality related traits (Lukaszewski & Roney, 2011).

In sum the literature shows convincing evidence for a relationship between self-rated trait measures (such as self-perceived physical attractiveness) and personality, but only mixed results or missing evidence for relationships between personality and other-rated or direct anthropometric measures of attractiveness or formidability (see Table 1 for an overview). Thus it could be possible that the heritability of personality variation is, contrary to the theory of facultative calibration, not reactive to heritability of other phenotypic traits, but solely related to how people *perceive* themselves in these traits. In the latter case the causal effect could go in the other direction, with individuals with certain personalities having a tendency to perceive themselves as more attractive or formidable, independent of their objective physique.

The aim of the present study was to probe the relationship between personality traits, formidability and physical attractiveness in an integrated way. In order to do so, we collected a broad range of formidability indicators and personality traits in two relatively large community samples from two countries, the UK and Germany.

Table 1

An Overview of main results in previous studies of facultative calibration

Women	Self-perceived attractiveness	Other-rated attractiveness	Physical strength	Anthropometric Measurements	References

	Self-perceived attractiveness	Other-rated attractiveness	Physical strength	Anthropometric Measurements	References
Anger	Positive association	Not tested	No association	No association	Price et al.; 2012, Sell et al., 2009
Extraversion	Positive association	Mixed Results	Mixed results	No association	Fink et al., 2016; Haysom, 2015; Lukaszewski, 2013; Lukaszewski & Roney, 2011, von Rueden et al., 2015
Narcissism	Positive association	Positive association	Not tested	Not tested	Gabriel et al., 1994; Holtzman & Strube, 2010
Sociosexual orientation	Unclear*	Unclear*	Unclear*	Not tested	Lukaszewski et al., 2014
Men	Self-perceived attractiveness	Other-rated attractiveness	Physical strength	Anthropometric Measurements	References
Anger proneness	Mixed Results	Not tested	Mixed results	Mixed results	Archer & Thanzami, 2007; Price et al., 2012; Sell et al.,

					200; Sell et al., 2016
Extraversion	Positive association	Mixed results	Positive association	No association	Fink et al., 2016; Haysom, 2015; Lukaszewski, 2013; Lukaszewski & Roney, 2011, von Rueden et al., 2015
Narcissism	Positive association	Positive association	Not tested	Not tested	Gabriel et al., 1994; Holtzman & Strube, 2010
Sociosexual orientation	Unclear*	Unclear*	Unclear*	Not tested	Lukaszewski et al., 2014

*No separate tests of significance for self-rated and objective trait measures

2. Study 1

2.1 Methods

2.1.1 Participants and Procedure

A total of 119 men and 124 women from Edinburgh (Scotland, UK) participated in the study (age 18-29 years, M = 21.5, SD = 2.1; years of completed education M = 15.6, SD = 3.5). Participants were either

undergraduate students recruited through the student subject pool of the University of Edinburgh and received course credit, or came from the local community, recruited via social networks and small advertisements, and received a compensation of £10. A total of 83.1% of the sample indicated their ethnicity as white, while the rest self-identified as Indian (2.9%), Chinese (5.3%), mixed (4.1%), or other (4.5%).

Laboratory assessments were conducted by same-sex experimenters. After signing an informed consent form, participants were seated upright in front of a 3DMD facial camera system with glasses and facial jewelry removed, asked to maintain a neutral expression, and a 3D picture of their faces was taken. Then they were asked to change into provided, tight fitting standardized underwear and scanned three times standing upright with a TC^2 NX-16 3D body scanner (Cary, NC, USA), following the procedure described in Price et al. (2012). Afterwards height and weight was measured with a statiometer and a digital scale, respectively. Flexed biceps circumference was measured with an anthropometric tape measure. Hand grip and upper body strength was measured with a dynamometer, and lung function with a spirometer. Finally, participants filled out a computerized battery of questionnaires (described below) in private. The study was approved by the University of Edinburgh Psychology Research Ethics Committee (application numbers 25-1112, 299-1112, 40-1213/2).

2.1.2 Measures

2.1.2.1 Personality Measures

The *anger* measurements were adopted from Sell et al. (2009), measured on a 5-point Likert scale from 1 = "strongly disagree" to 5 = "strongly agree", and included *Proneness to Anger (*11 items, Cronbach's α=0.77), *Utility of Personal Aggression (*16 items, Cronbach's α=0.77), *Success in Conflict (*7 items, Cronbach's α=0.81), and *History of Fighting* (5 items, Cronbach's α=0.73). Typical items for each scale were as follows: "It is harder to get me angry than other people" (*Proneness to Anger*, reverse coded), "If I don't respond to provocations and do something to make the wrong-doers pay, they'll just do more to hurt me in the future" (*Utility of Personal Aggression*), "When there's a dispute, I usually get my way" (*Success in Conflict*), "I have physically intimidated someone who had it coming" (*History of Fighting*).

Vengefulness (4 items, 5-point Likert scale, 1 = "strongly disagree" to 5 = "strongly agree", Cronbach's α=0.72) was measured using the Revenge subscale of the Transgression-Related Interpersonal Motivations (TRIM) inventory (McCullough et al., 1998). A typical item was: "When someone angers me or hurts my feelings, I usually find a way to make this person regret it."

Extraversion (48 items, 5-point Likert scale, 1 = "strongly disagree" to 5 = "strongly agree", Cronbach's α=0.91) was measured using the NEO-PI-R Extraversion scale (Costa & McCrae, 1992), including the six facets Warmth (Cronbach's α=0.81), Gregariousness (Cronbach's α=0.81), Assertiveness (Cronbach's α=0.81), Activity (Cronbach's α=0.72), Excitement Seeking (Cronbach's α=0.65), and Positive Emotion (Cronbach's α=0.81). All facets consisted of 8 items. Typical items for each facet were: "I really enjoy talking to people." (Warmth), "I like to have a lot of people around me." (Gregariousness), "I am dominant, forceful, and assertive." (Assertiveness), "I often feel as if I'm bursting with energy." (Activity), "I like to be where the action is." (Excitement Seeking), "I am a cheerful, high-spirited person." (Positive Emotion).

Dominance (11 items, 5-point Likert scale, 1 = "strongly disagree" to 5 = "strongly agree", Cronbach's α=0.82) was measured using the Interpersonal Personality Item Pool version of the dominance subscale from the CPI narcissism scale (Goldberg et al., 2006; Gough, 1956). A typical item was: "I impose my will on others."

Shyness (5 items, 5-point Likert scale, agreement format from 1 = "not at all" to 5 = "completely", Cronbach's α=0.84) was measured using the five item Shyness Scale (Asendorpf & Wilpers, 1998). A typical item was: "I feel inhibited when I am with other people".

Narcissism was measured using the Narcissistic Admiration and Rivalry Questionnaire (NARQ; Back et al., 2013), comprising the two dimensions Admiration (9 items, 5-point Likert scale, 1 = "not agree at all" to 6 = "agree completely", Cronbach's α=0.77) and Rivalry (9 items, 5-point Likert scale, Cronbach's α=0.73). Typical items were: "I will someday be famous" (Admiration) and "I react annoyed if another person steals the show from me." (Rivalry).

Sociosexual Orientation (9 items, 5-point response scales, Cronbach's α=0.88) was measured using the revised Sociosexual Orientation Inventory (SOI-R; Penke & Asendorpf, 2008), comprising the three facets Attitude (3 items, Cronbach's α=0.88), Behavior (3 items, Cronbach's α=0.85), and Desire (3 items, Cronbach's α=0.83). Typical items were: "With how many different partners have you had sex within the past 12 months?" (Behavior), "Sex without love is OK." (Attitude), "In everyday life, how often do you have spontaneous fantasies about having sex with someone you have just met?" (Desire).

2.1.2.2 Measures of Physical Attractiveness

Self-perceived physical attractiveness was measured as a 3-item aggregate (Cronbach's α=0.82) based on questions similar to those used in previous studies (cf. Lukaszewski & Roney, 2011; Price et al., 2012; Sell et al., 2009): "I am more attractive than __% of others of my sex"; "On a scale from 1 to 10, how physically attractive are you?", "Compared to others I'm a very attractive person" (on a Likert scale from 1="strongly disagree" to 5="strongly agree").

Since a tighter link in a attractiveness-personality-relationship could be expected for measures that are more directly related to mating success, we computed *self-perceived mating success* as a 3-item aggregate (Cronbach's α=0.82) based on 7-point Likert scale items (1= "not at all" to 7 = "very") from the self-perceived Mate Value Scale MVS (Landolt, Lalumière, & Quinsey, 1995): "Members of the opposite sex are attracted to me.", "Members of the opposite sex notice me.", "I do not receive many compliments from members of the opposite sex.".

Other-rated physical attractiveness was judged by eight male and eight female raters, mostly undergraduate students of the University of Edinburgh (age *M* = 22.1 years, *SD* = 1.1). Raters saw rotating animations ('beauty turns') of the body scans, with heads removed to focus attention on body attractiveness and uniform grey color (so free of skin color cues; similar to Smith et al., 2007). Relative height differences of the stimulus subjects were maintained in the presentations. Beauty turns were displayed on a computer screen using the Eprime software, and evaluated them individually on a (7-point Likert scale from 1= "not attractive" to 7 = "very attractive", interrater agreement Cronbach's α=0.92).

Residual self-perceived physical attractiveness was computed as the residuals from a regression of self-perceived on other-rated physical attractiveness. This additional index of self-perceived physical attractiveness is controlled for the consensual, arguably more objective outside perception of one's body attractiveness, thereby getting closer the subjective component of self-perceived attractiveness.

2.1.2.3 Formidability Measures

Strength was assessed as hand grip strength for both hands and upper body strength, all measured with a Saehan SH500 dynamometer following the procedure described in the Appendix of Sell et al. (2009). Handedness was assessed using the Edinburgh Handedness Inventory (Oldfield, 1971). Each strength measure was taken three times and the maximum performance was used for further analyses. An overall strength variable was computed as the mean of z-standardized measures of dominant hand grip and upper body strength (which were highly correlated; r=0.83, p<0.001).

Lung function was measured with a spirometer (MicroPlus, CareFusion). Lung forced vital capacity (FVC) and forced expiratory volume per 1 second (FEV) were assessed three times, and the maximum performance values for each measure were z-standardized and averaged into an overall lung function variable. We included lung function as an additional measure of formidability indicating a person's aerobic fitness and therefore potentially physical competitiveness.

Body masculinity was calculated, following Price et al. (2012), as the regression score of the first unrotated principal component extracted across sexes from the following sexually dimorphic variables (effect sizes of group differences between sexes are given in parentheses): height (*d*=2.17), dominant arm flexed biceps circumference (*d*=1.84), and indices calculated from automatic measurements taken by the TC^2 NX-16 body scanner software (average of measures from three scans), including shoulder breadth (*d*=1.69), forearm circumference (*d*=1.73), chest circumference (*d*=1.65), waist-to-hip ratio (*d*=1.34), bust-to-underbust ratio (*d*=-2.80), and leg-to-body ratio (*d*=-1.07). The PCA explained 63.48% of the variables` variance.

BMI-controlled body masculinity was computed as the residuals from a regression of *Body masculinity* on BMI. This measure was intended as a robustness check to separate the index of *Body masculinity* from the

Page 145

influence of the BMI, approximatively controlling for differences in body measurements that are not due to sexual dimorphisms in body shape or muscularity but due to idiosyncratic differences in body fat (although BMI is correlated to muscularity as well; Heymsfield, Scherzer, Pietrobelli, Lewis, & Grunfeld, 2009).

Facial masculinity was computed sensu Penton-Voak et al. (2001) from 14 landmarks placed on the 3D facial photographs using Morphanalyser (Tiddeman, Duffy, & Rabey, 2000). For the facial masculinity index, z-standardized measures of face width to lower face height, eye size, and cheekbone prominence were subtracted from the z-standardized ratio of lower face to face height.

2.1.3 Statistical Analyses

We z-standardized all variables (except for the dichotomous control variable ethnicity, white vs. non-white) prior to analysis and graphically inspected the data for normality and outliers. From the original sample (119 males, 124 females) we excluded 3 influential cases (2 males, 1 female) based on outliers in body appearance (values beyond 1.5 * interquartile range above or below the third or first quartile, respectively) that showed high leverage in regression diagnostics (graphical inspection of the bivariate distribution between residuals and leverage of data points as well as distribution of Cook's distances; see figures S1.1 and S1.2 in the supplementary material). For the resulting sample (117 males, 123 females) we computed zero-order Pearson correlation coefficients between all variables and their bootstrapped 95% confidence intervals. To account for multiple testing we also adjusted all p-values of the correlations in the result section for false discovery rate (FDR; Benjamini & Hochberg, 1995), i.e. controlled for the expected proportion of falsely rejected hypotheses among all rejected hypotheses, using the R package "psych" (Revelle, 2016). As a robustness analysis we additionally computed partial Pearson correlation coefficients controlling for age (since age was an important influence on the results of Price et al., 2012) and ethnicity of participants (separating between white and non-white). Then we computed the congruency coefficient Rc (Abdi, 2010) between the correlation matrices with and without control variables, where a value of Rc=1 would indicate complete congruency between both matrices, and checked whether changes in statistical significance (in terms of a threshold $p < 0.05$) occurred.

2.2 Results

First, we assessed how self-perceived attractiveness and self-perceived mating success were related to each other, as well as to objective body measures and other-rated attractiveness. Whereas self-rated attractiveness and mating success were strongly correlated in men (r=0.69, p<0.001) and women (r=0.61, p<0.001), self-perceived attractiveness and mating success correlated with objective indicators of physical appearance only among men (other-rated attractiveness, strength, and height; see Table 2).

Self-perceived attractiveness was, as expected, positively correlated with the personality measures Extraversion, Dominance (only in men), and Narcissistic Admiration, as well as negatively with Shyness. Furthermore it showed a positive relationship with Sociosexual Behavior and also partly with Anger or Aggressiveness in both sexes. Other-rated attractiveness was overall not correlated with any personality measures (except for Success in Conflict in women; see Table 3). For men we even found a negative correlation between other-rated attractiveness and Proneness to Anger, which was contrary to the expected direction (Table 3).

Extraversion was completely unrelated to objective measures of formidability such as physical masculinity and strength on the domain level. However, in the male subsample, the facets Assertiveness and Activity positively correlated with body masculinity and physical strength, as well as height with Gregariousness (see Table 3). With one exception (lung function positively related to Excitement Seeking) we did not find any such correlations for women. Sociosexual Orientation and Anger Proneness were also largely unrelated to objective body measures in both sexes, however with a few exceptions: Positive correlations occurred between body masculinity, upper body strength and Utility of Personal Aggression, between physical strength and Narcissistic Admiration, and between height and short-term mating behavior in men. In addition, we found an unpredicted negative correlation between lung function and Narcissistic Rivalry in men.

The similarity of partial correlations controlling for age and ethnicity with zero-order correlations was highly significant (men: Rc=0.999; 95% CI= [0.999;0.999]; p<0.001; women: Rc=0.998; 95% CI= [0.996;0.999]; p<0.001). Changes of statistical significance between partial- and zero-order correlations occurred in six cases for the male and five cases for the female sample. All changes, except the correlation

among BMI-controlled Body Masculinity and History of Fighting in men, reflected correlations either among two personality variables or two formidability variables and did thus not affect the interpretation of relationships subject to facultative calibration. The results of correlations based on single anthropometric measures of the body and the face of individuals (which were combined into measures of body masculinity and facial masculinity in Table 3) can be found in the supplementary material (S3).

Table 2

Pearson correlation coefficients between self-rated and objective formidability measures

Women	SP attractiveness	SP mating success
Other-rated attractiveness	0.14 [-0.09;0.34]	0.16 [-0.03;0.35]
Body masculinity	-0.15 [-0.30;0.03]	-0.16 [-0.28;-0.02]
BMI-controlled body masculinity	0.03 [-0.12;0.22]	-0.03 [-0.16;0.12]
Upper body size	-0.13 [-0.29;0.06]	-0.14 [-0.26;0.03]
Facial masculinity	-0.16 [-0.34;0.02]	-0.12 [-0.29;0.05]
Strength	0.06 [-0.15;0.21]	0.08 [-0.09;0.23]
Lung function	-0.01 [-0.16;0.13]	-0.02 [-0.18;0.13]
Height	-0.02 [-0.16;0.14]	0.01 [-0.12;0.19]
BMI	*-0.23 [-0.43;-0.01]*	-0.19 [-0.34;0.00]
Men	SP attractiveness	SP mating success
Other-rated attractiveness	***0.33 [0.16;0.52]***	***0.23 [0.07;0.41]***
Body masculinity	0.14 [-0.02;0.34]	0.02 [-0.14;0.24]
BMI-controlled	***0.30 [0.17;0.44]***	*0.17 [0.03;0.35]*

body masculinity

Upper body size	0.13 [-0.06;0.36]	-0.02 [-0.2;0.20]
Facial masculinity	0.05 [-0.15;0.19]	-0.03 [-0.18;0.12]
Strength	*0.26 [0.10;0.41]*	0.12 [-0.05;0.29]
Lung function	0.10 [-0.02;0.27]	0.10 [-0.05;0.25]
Height	*0.26 [0.12;0.44]*	*0.25 [0.08;0.40]*
BMI	-0.06 [-0.28;0.16]	-0.11 [-0.29;0.12]

Significant results in terms of FDR-adjusted p-values <.05 are displayed in bold, confidence intervals not containing the value 0 in italics.

Table 3

Pearson correlation coefficients between formidability and personality measures from the Study 1, female subsample

	Anger					Extraversion							Additional Personality				Sociosexual	
	UA	AP	SC	FH	V	E	E1	E2	E3	E4	E5	E6	D	S	NR	NA	SOI-R	S-A
Self-perceived attractiveness	-0.06 [-0.28; 0.11]	-0.04 [-0.26; 0.16]	**0.30 [0.13; 0.47]**	0.17 [-0.02; 0.32]	0.08 [-0.10; 0.25]	**0.35 [0.23;0.50]**	**0.24 [0.06; 0.42]**	**0.26 [0.08; 0.42]**	0.17 [0.00; 0.34]	*0.19 [0.04; 0.36]*	**0.25 [0.11; 0.38]**	**0.30 [0.16; 0.45]**	0.13 [-0.11; 0.33]	**-0.35 [-0.51; -0.21]**	0.06 [-0.15; 0.24]	**0.58 [0.46; 0.69]**	*0.17 [0.03; 0.35]*	0.07 [-0.09; 0.26]
Residual self-perceived attractiveness	-0.07 [-0.28; 0.09]	-0.06 [-0.29; 0.15]	*0.21 [0.03; 0.42]*	0.17 [-0.01; 0.33]	0.05 [-0.15; 0.23]	**0.31 [0.15;0.47]**	*0.21 [0.00; 0.41]*	**0.25 [0.10; 0.41]**	0.14 [-0.05; 0.36]	0.12 [-0.04; 0.29]	**0.29 [0.15; 0.44]**	**0.26 [0.09; 0.44]**	0.10 [-0.15; 0.32]	**-0.33 [-0.5; -0.18]**	0.05 [-0.15; 0.21]	**0.53 [0.41; 0.67]**	*0.18 [0.02; 0.36]*	0.05 [-0.13; 0.26]
Self-perceived mating success	-0.05 [-0.23; 0.10]	-0.08 [-0.28; 0.14]	**0.38 [0.21; 0.53]**	0.08[-0.10; 0.24]	0.03 [-0.16; 0.20]	**0.38 [0.23;0.51]**	**0.28 [0.07; 0.49]**	**0.30 [0.13; 0.44]**	0.19 [-0.05; 0.34]	*0.20 [0.04; 0.36]*	**0.20 [0.04; 0.35]**	**0.39 [0.26; 0.54]**	0.10 [-0.17; 0.25]	**-0.40 [-0.52; -0.24]**	-0.11 [-0.29; 0.05]	**0.50 [0.37; 0.61]**	0.14 [-0.02; 0.32]	0.10 [-0.07; 0.27]
Other-rated attractiveness	0.11 [-0.04; 0.22]	0.10 [-0.08; 0.26]	**0.21 [0.05; 0.35]**	-0.04 [-0.23; 0.11]	*0.17 [0.01;0.31]*	0.01 [-0.19; 0.19]	-0.05 [-0.20; 0.10]	-0.08 [-0.27; 0.13]	0.02 [-0.16; 0.17]	0.14 [-0.04; 0.33]	0.00 [-0.19; 0.22]	0.02 [-0.15; 0.19]	0.13 [-0.03; 0.28]	0.06 [-0.09; 0.24]	0.09 [-0.10; 0.25]	0.08 [-0.08; 0.22]	-0.10 [-0.25; 0.08]	-0.05 [-0.21; 0.10]
Body masculinity	-0.09 [-0.21; 0.04]	-0.08 [-0.22; 0.09]	-0.05 [-0.21; 0.10]	0.00 [-0.17; 0.15]	-0.13 [-0.29; 0.05]	0.09 [-0.05; 0.26]	0.04 [-0.08; 0.18]	0.01 [-0.14; 0.19]	0.13 [-0.03; 0.29]	0.03 [-0.12; 0.21]	0.06 [-0.10; 0.25]	0.09 [-0.04; 0.26]	-0.01 [-0.16; 0.13]	-0.06 [-0.20; 0.08]	-0.17 [-0.33; 0.03]	-0.10 [-0.27; 0.11]	0.11 [-0.02; 0.27]	0.08 [-0.05; 0.21]
BMI-controlled body masculinity	-0.09 [-0.20; 0.05]	-0.02 [-0.22, 0.15]	0.03 [-0.14; 0.16]	-0.01 [-0.17; 0.15]	-0.11 [-0.23; 0.04]	0.10 [-0.03; 0.27]	-0.04 [-0.15; 0.13]	-0.02 [-0.18; 0.13]	0.10 [-0.07; 0.26]	*0.14 [0.01; 0.31]*	0.15 [-0.03; 0.31]	0.08 [-0.10; 0.26]	0.10 [-0.06; 0.22]	-0.01 [-0.16; 0.12]	-0.09 [-0.22, 0.06]	0.02 [-0.13; 0.19]	0.06 [-0.08; 0.24]	0.06 [-0.10; 0.27]
Upper body size	-0.08 [-0.20; 0.06]	-0.12 [-0.29; 0.03]	-0.04 [-0.22; 0.12]	0.02 [-0.14; 0.18]	-0.09 [-0.26; 0.05]	0.03 [-0.10; 0.21]	0.02 [-0.10; 0.17]	0.00 [-0.12; 0.17]	0.06 [-0.09; 0.24]	-0.04 [-0.19; 0.15]	0.01 [-0.16; 0.19]	0.07 [-0.10; 0.25]	-0.07 [-0.21; 0.10]	-0.04 [-0.17; 0.09]	-0.12 [-0.28; 0.03]	-0.11 [-0.29; 0.13]	*0.17 [0.01; 0.32]*	0.12 [0.00; 0.28]
Facial masculinity	*0.23 [0.03; 0.39]*	0.02 [-0.17; 0.2]	0.07 [-0.16; 0.26]	0.01 [-0.19; 0.15]	-0.04 [-0.22; 0.18]	0.06 [-0.17; 0.20]	-0.10 [-0.26; 0.07]	0.02 [-0.14; 0.19]	0.01 [-0.15; 0.14]	0.20 [-0.04; 0.36]	0.12 [-0.10; 0.25]	-0.02 [-0.21; 0.13]	0.09 [-0.12; 0.27]	-0.15 [-0.30; 0.02]	*0.13 [0.01;0.29]*	0.02 [-0.17; 0.16]	-0.02 [-0.16; 0.12]	0.00 [-0.16; 0.17]
Strength	*-0.17*	-0.17	0.02	-0.10	*-0.18*	0.13	0.06	0.00	-0.03	*0.19*	0.14	*0.18*	-0.07	-0.05	*-0.14*	-0.01	0.00	0.02

[-0.34; -0.02]	[-0.34; 0.06]	[-0.12; 0.19]	[-0.33; 0.10]	*[-0.33; -0.02]*	[-0.03; 0.28]	[-0.06; 0.23]	[-0.16; 0.20]	[-0.18; 0.11]	*[0.07; 0.35]*	[-0.05; 0.30]	*[0.01; 0.33]*	[-0.18; 0.10]	[-0.20; 0.12]	*[-0.30; -0.01]*	[-0.16; 0.16]	[-0.14; 0.13]	[-0.15; 0.22]
Lung function -0.16 [-0.32; 0.02]	-0.08 [-0.25; 0.11]	-0.03 [-0.20; 0.13]	-0.13 [-0.34; 0.04]	-0.08 [-0.3; 0.07]	0.10 [-0.08; 0.26]	-0.04 [-0.20; 0.17]	0.04 [-0.11; 0.20]	-0.04 [-0.19; 0.13]	0.13 [-0.07; 0.31]	**0.25** *[0.13; 0.39]*	0.06 [-0.17; 0.20]	0.10 [-0.09; 0.33]	0.02 [-0.13; 0.18]	0.05 [-0.11; 0.23]	-0.02 [-0.14; 0.14]	*0.16* *[0.01; 0.35]*	0.09 [-0.08; 0.28]
Height -0.21 [-0.36; -0.02]	-0.14 [-0.34; 0.06]	-0.06 [-0.17; 0.15]	*-0.16* *[-0.34; -0.01]*	-0.07 [-0.23; 0.10]	0.00 [-0.14; 0.17]	-0.07 [-0.21; 0.13]	-0.05 [-0.20; 0.12]	0.03 [-0.11; 0.25]	0.01 [-0.12; 0.16]	-0.01 [-0.16; 0.17]	0.06 [-0.08; 0.21]	-0.01 [-0.20; 0.2]	0.09 [-0.10; 0.21]	-0.04 [-0.19; 0.12]	-0.07 [-0.19; 0.13]	0.02 [-0.14; 0.16]	-0.02 [-0.14; 0.14]
BMI -0.03 [-0.17; 0.10]	-0.10 [-0.27; 0.07]	-0.10 [-0.27; 0.08]	0.01 [-0.13; 0.18]	-0.09 [-0.26; 0.07]	0.02 [-0.14; 0.19]	0.08 [-0.03; 0.21]	0.03 [-0.13; 0.20]	0.08 [-0.10; 0.22]	-0.10 [-0.27; 0.08]	-0.07 [-0.28; 0.10]	0.04 [-0.13; 0.25]	-0.12 [-0.26; 0.04]	-0.06 [-0.21; 0.05]	*-0.15* *[-0.26; -0.01]*	-0.16 [-0.36; 0.07]	0.09 [-0.07; 0.26]	0.04 [-0.11; 0.20]

Significant results in terms of FDR-adjusted p-values <.05 are displayed in bold, confidence intervals not containing the value 0 in italics. UA= Utility of Personal Aggression, AP= Proneness to Anger, SC= Success in Conflict, FH= History of Fighting, V= Vengefulness, E= Extraversion, E1= Warmth, E2= Gregariousness, E3= Assertiveness, E4= Activity, E5= Excitement-Seeking, E6= Positive Emotions, D= Dominance, S= Shyness, NR= Narcissistic Rivalry, NA= Narcissistic Admiration, SOI-R= revised Sociosexual Orientation Inventory, S-A= Sociosexual Attitudes, S-B= Sociosexual Behavior, S-D= Sociosexual Desire

Table 4

Pearson correlation coefficients between formidability and personality measures from the Study 1, male subsample

	Anger					Extraversion								Additional Personality				Sociosexual Orie
	UA	AP	SC	FH	V	E	E1	E2	E3	E4	E5	E6	D	S	NR	NA	SOI-R	S-A	S-
Self-perceived attractiveness	**0.36** **[0.23; 0.50]**	*0.20* *[0.06; 0.38]*	**0.58** **[0.45;0.71]**	**0.39** **[0.24; 0.53]**	**0.24** **[0.09; 0.38]**	**0.46** **[0.36; 0.56]**	*0.17* *[0.05; 0.30]*	**0.38** **[0.23; 0.55]**	**0.46** **[0.28; 0.59]**	**0.50** **[0.39; 0.60]**	**0.31** **[0.14; 0.46]**	**0.30** **[0.14; 0.46]**	**0.24** **[0.12; 0.38]**	**-0.45** **[-0.57; -0.33]**	-0.03 [-0.17; 0.14]	**0.52** **[0.39; 0.61]**	*0.18* *[0.05; 0.33]*	0.16 [0.00; 0.35]	
Residual self-perceived attractiveness	**0.40** **[0.29; 0.53]**	**0.27** **[0.13; 0.43]**	**0.56** **[0.43;0.69]**	**0.45** **[0.32;0.59]**	**0.30** **[0.14; 0.44]**	**0.44** **[0.32; 0.55]**	*0.19* *[0.07; 0.34]*	**0.38** **[0.23; 0.57]**	**0.43** **[0.28; 0.54]**	**0.46** **[0.31; 0.57]**	**0.28** **[0.11; 0.47]**	**0.27** **[0.11; 0.46]**	**0.28** **[0.14; 0.42]**	**-0.42** **[-0.56;-0.31]**	0.01 [-0.18; 0.21]	**0.48** **[0.35; 0.58]**	*0.18* *[0.03; 0.34]*	0.14 [-0.03; 0.35]	
Self-perceived mating success	*0.18* *[0.03; 0.37]*	0.04 [-0.14; 0.27]	**0.43** **[0.24;0.59]**	**0.27** **[0.10; 0.45]**	*0.18* *[0.04; 0.34]*	**0.40** **[0.27; 0.57]**	**0.22** **[0.07; 0.38]**	**0.37** **[0.22; 0.54]**	**0.27** **[0.07; 0.45]**	**0.36** **[0.19; 0.56]**	**0.32** **[0.14; 0.51]**	**0.29** **[0.07; 0.46]**	*0.18* *[0.03; 0.35]*	**-0.38** **[-0.56; -0.23]**	-0.08 [-0.25; 0.14]	**0.37** **[0.23; 0.52]**	**0.32** **[0.20; 0.47]**	**0.33** **[0.18; 0.50]**	

151

Other-rated attractiveness	-0.01 [-0.16; 0.17]	**-0.22 [-0.37; -0.07]**	0.12 [-0.06; 0.3]	-0.07 [-0.22; 0.08]	-0.19 [-0.39; 0.00]	0.11 [-0.06; 0.27]	-0.08 [-0.24; 0.07]	0.07 [-0.10; 0.24]	0.09 [-0.11; 0.26]	*0.17 [0.02; 0.36]*	0.12 [-0.05; 0.28]	0.15 [0.00; 0.32]	-0.09 [-0.25; 0.07]	-0.07 [-0.24; 0.09]	-0.18 [-0.34; 0.02]	0.14 [-0.01; 0.31]	0.09 [-0.11; 0.23]	0.11 [-0.07; 0.27]	0. [- 0…
Body masculinity	**0.27 [0.13; 0.41]**	0.04 [-0.15; 0.19]	0.15 [-0.05; 0.31]	0.19 [-0.01; 0.31]	0.10 [-0.07; 0.26]	0.14 [-0.01; 0.31]	0.08 [-0.06; 0.24]	-0.01 [-0.12; 0.18]	**0.25 [0.04; 0.42]**	**0.24 [0.07; 0.42]**	0.05 [-0.08; 0.22]	0.03 [-0.16; 0.21]	0.13 [-0.08; 0.28]	0.00 [-0.20; 0.15]	-0.01 [-0.20; 0.15]	0.11 [-0.15; 0.31]	0.10 [-0.06; 0.24]	0.02 [-0.14; 0.2]	…
BMI-controlled body masculinity	**0.23 [0.05; 0.37]**	-0.03 [-0.19, 0.14]	0.17 [-0.02; 0.41]	*0.18 [0.04; 0.31]*	0.05 [-0.10; 0.21]	*0.16 [0.02; 0.32]*	0.01 [-0.18; 0.20]	0.05 [-0.10; 0.21]	**024 [0.12; 0.38]**	**0.26 [0.09; 0.41]**	0.10 [-0.03; 0.29]	0.08 [-0.09; 0.29]	0.04 [-0.16; 0.20]	-0.04 [-0.22; 0.11]	-0.15 [-0.32; 0.05]	0.14 [-0.03; 0.28]	0.14 [-0.02; 0.32]	0.10 [-0.06; 0.24]	0. […]
Upper body size	**0.23 [0.08; 0.36]**	0.04 [-0.16; 0.22]	0.15 [-0.05; 0.30]	0.13 [-0.08; 0.30]	0.10 [-0.04; 0.24]	0.14 [-0.02; 0.33]	0.13 [-0.01; 0.31]	-0.02 [-0.16; 0.20]	**0.24 [0.04; 0.42]**	*0.20 [0.01; 0.39]*	0.05 [-0.08; 0.22]	0.05 [-0.13; 0.22]	0.13 [-0.06; 0.27]	-0.01 [-0.19; 0.14]	0.00 [-0.15; 0.14]	0.09 [-0.16; 0.27]	0.00 [-0.14; 0.15]	-0.08 [-0.21; 0.09]	…
Facial masculinity	0.06 [-0.08; 0.22]	0.05 [-0.12; 0.26]	0.04 [-0.14; 0.28]	0.10 [-0.07; 0.29]	-0.06 [-0.22; 0.10]	0.06 [-0.18; 0.33]	0.11 [-0.11; 0.36]	-0.01 [-0.21; 0.18]	0.04 [-0.12; 0.23]	0.01 [-0.19; 0.22]	0.12 [-0.04; 0.37]	0.02 [-0.20; 0.24]	-0.01 [-0.19; 0.20]	-0.05 [-0.26; 0.15]	0.07 [-0.13; 0.24]	0.09 [-0.07; 0.27]	-0.02 [-0.22; 0.18]	-0.12 [-0.27; 0.06]	…
Strength	0.09 [-0.11; 0.26]	-0.12 [-0.37; 0.06]	0.20 [-0.04; 0.39]	0.09 [-0.08; 0.29]	-0.08 [-0.24; 0.10]	*0.18 [0.02; 0.34]*	0.08 [-0.07; 0.23]	0.02 [-0.11; 0.20]	**0.28 [0.13; 0.44]**	**0.26 [0.08; 0.44]**	0.08 [-0.06; 0.26]	0.12 [-0.04; 0.31]	0.10 [-0.05; 0.25]	-0.12 [-0.30; 0.03]	0.04 [-0.11; 0.22]	**0.28 [0.15; 0.43]**	0.01 [-0.17; 0.18]	0.01 [-0.15; 0.21]	…
Lung function	0.12 [-0.02; 0.27]	-0.09 [-0.25; 0.07]	0.05 [-0.11; 0.20]	0.13 [-0.05; 0.27]	-0.08 [-0.23; 0.09]	0.06 [-0.08; 0.21]	-0.03 [-0.18; 0.14]	0.04 [-0.07; 0.17]	0.08 [-0.05; 0.22]	0.13 [-0.02; 0.28]	0.02 [-0.16; 0.17]	0.02 [-0.11; 0.15]	-0.01 [-0.18; 0.11]	0.02 [-0.13; 0.15]	**-0.26 [-0.38; -0.08]**	-0.08 [-0.2; 0.08]	0.05 [-0.13; 0.25]	0.01 [-0.16; 0.18]	…
Height	0.06 [-0.12; 0.26]	-0.11 [-0.27; 0.05]	0.10 [-0.08; 0.30]	*0.16 [0.04; 0.31]*	0.04 [-0.15; 0.22]	*0.20 [0.05; 0.35]*	0.07 [-0.03; 0.26]	**0.25 [0.09; 0.41]**	0.10 [-0.05; 0.28]	*0.18 [0.02; 0.38]*	0.16 [-0.02; 0.34]	0.13 [-0.02; 0.31]	-0.04 [-0.19; 0.16]	-0.17 [-0.33; -0.02]	-0.16 [-0.30; 0.05]	0.03 [-0.10; 0.25]	*0.15 [0.02; 0.29]*	0.04 [-0.12; 0.19]	**0. […]**
BMI	*0.19 [0.06; 0.33]*	0.08 [-0.08; 0.24]	0.07 [-0.12; 0.24]	0.12 [-0.10; 0.29]	0.10 [-0.04; 0.26]	0.06 [-0.08; 0.24]	0.10 [-0.04; 0.26]	-0.05 [-0.17; 0.2]	0.15 [-0.03; 0.34]	0.13 [-0.03; 0.3]	-0.01 [-0.13; 0.17]	-0.03 [-0.21; 0.15]	0.16 [0.00; 0.28]	0.03 [-0.17; 0.19]	0.11 [-0.04; 0.27]	0.04 [-0.22; 0.23]	0.02 [-0.14; 0.18]	-0.05 [-0.20; 0.11]	…

152

Significant results in terms of FDR-adjusted p-values <.05 are displayed in bold, confidence intervals not containing the value 0 in italics. UA= Utility of Personal Aggression, AP= Proneness to Anger, SC= Success in Conflict, FH= History of Fighting, V= Vengefulness, E= Extraversion, E1= Warmth, E2= Gregariousness, E3= Assertiveness, E4= Activity, E5= Excitement-Seeking, E6= Positive Emotions, D= Dominance, S= Shyness, NR= Narcissistic Rivalry, NA= Narcissistic Admiration, SOI-R= revised Sociosexual Orientation Inventory, S-A= Sociosexual Attitudes, S-B= Sociosexual Behavior, S-D= Sociosexual Desire

3. Study 2

3.1 Methods

Note that since Study 2 was overall very similar to Study 1, we do only describe those methodological aspects of Study 2 in detail that differed from Study 1.

3.1.1 Participants and Procedure

165 men (age: M=24.3, SD=3.2 years) from the local community of Göttingen (Germany), mostly university undergraduates, were preselected for being heterosexual (to satisfy demands of another study based on these data; heterosexual orientation on a 7-point Kinsey scale; Kinsey et al., 1948) and recruited via an online database, in exchange for monetary compensation. The participants were asked to fill out a battery of questionnaires (described below) and subsequently were body-scanned using a Vitus^smartXXL 3D bodyscanner (Human Solutions GmbH, Kaiserslautern, Germany) and a 3dMD face scanner. All participants were scanned three times, while only wearing standardized tight underwear in the size of their choice (small to extra-extra-large). Participants were instructed to stand still in a standardized posture (standing upright with legs hip-widely apart, arms stretched out and held slightly away from the body, hands making a fist with thumbs showing forward, and head positioned in accordance with the Frankfort Horizontal) and breathe normally during the scanning process (ca. 10 secs. each). Additionally, body height (in cm) was measured twice using a statiometer. The participants stood upright without wearing shoes; the two values were averaged. Weight (in kg) was measured as part of each body scanning process with the integrated scale SECA 635 (SECA, Hamburg, Germany); the three values were averaged.

3.1.2 Measures

3.1.2.1 Personality Measures

Extraversion (8 items, Cronbach's α=0.87) was measured with the German version of the Big Five Inventory (BFI; Lang et al., 2001; 5-point Likert scale from 1 = "strongly disagree" to 5 = "strongly agree"). A typical items was: "I see myself as someone who is talkative".

In an exploratory manner we also included the remaining Big Five personality domains *Neuroticism* (7 items, Cronbach's α=0.81)*, Openness to Experience* (10 items, Cronbach's α=0.78)*, Agreeableness* (8 items, Cronbach's α=0.73)*,* and *Conscientiousness* (9 items, Cronbach's α=0.84)*,* which were measured with the German version of the Big Five Inventory as well (5-point Likert scale from 1 = "strongly disagree" to 5 = "strongly agree"). Typical items were as follows: "I see myself as someone who: is depressed, blue" (Neuroticism); "is original, comes up with new ideas" (Openness to Experience); "has a forgiving nature" (Agreeableness); "does a thorough job" (Conscientiousness).

Dominance was measured using five out of the eight items of the Interpersonal Adjective List (Jacobs & Scholl, 2005) that assess the PA facet (dominance-assured) of the Interpersonal Circumplex (Wiggins et al., 1988; 8-point Likert scale, 1= "extremely inaccurate" to 8 = "extremely accurate", Cronbach's α=0.74). A typical item was: "I am forceful."

Shyness was measured using five out of the eight items of the Interpersonal Adjective List (Jacobs & Scholl, 2005) that assess the HI facet (unassured-submissive; 8-point Likert scale, 1= "extremely inaccurate" to 8 = "extremely accurate", Cronbach's α=0.75). A typical item was: "I am timid."

Narcissism was measured using the short version of the Narcissistic Admiration and Rivalry Questionnaire (NARQ; Back et al., 2013), comprising the two dimensions Admiration (3 items, 6-point Likert scale, Cronbach's α=0.71) and Rivalry (3 items, 6-point Likert scale, Cronbach's α=0.56). For typical items see methods of Study 1.

Sociosexual Orientation was measured using the revised Sociosexual Orientation Inventory (SOI-R; Penke & Asendorpf, 2008) on a 9-point scale comprising the three facets Attitude (3 items, Cronbach's α=0.71), Behavior (3 items, Cronbach's α=0.86), and Desire (3 items, Cronbach's α=0.85). For typical items see methods of Study 1.

Anger Proneness and *Vengefulness* were not included in Study 2.

3.1.2.2 Measures of Physical Attractiveness

Self-perceived physical attractiveness was based on an aggregate of three items (attractive, sexy, appealing) rated on a 5-point Likert scale from 1 = "very unattractive" to 5 = "very attractive" (Cronbach's α=0.85).

Other-rated body attractiveness was rated by 31 females (the stimuli sample was divided into two sets in order to avoid rater fatigue effects, we had 15 and 16 raters for sets 1 and 2, respectively) on an 11-point scale (from -5 = "very unattractive" to +5 = "very attractive") in response to the question "How attractive is this man?" (set 1: α= 0.93, set 2: α= 0.94). Animated videos of a body scan turning around its vertical axis (similar to Smith et al., 2007) were created using AnthroScan software ("beauty turns", duration: 8 sec. each; Human Solutions GmbH, Kaiserslautern, Germany) and used as stimuli.

Other-rated facial attractiveness was based on facial photographs of the study`s participants and rated by 12 independent female raters on an 11-point scale (from -5 = "very unattractive" to +5 = "very attractive") and included as the mean of the responses to the two questions "How sexually attractive do you find this man?" (α=0.86) and "How attractive is this man for a long-term relationship?" (α=0.85).

Residual self-perceived physical attractiveness was calculated as in Study 1, but divided into two measures, the first representing residuals resulting from a regression of self-perceived attractiveness on other-rated body attractiveness, the second being residuals from a regression on other-rated facial attractiveness.

Self-perceived mating success was measured exactly as in Study 1.

3.1.2.3 Formidability Measures

Strength was assessed as in Study 1, but testing hand grip strength for participants' self-reported dominant hand only (88.2% used their right, the remaining 11.8% their left hand). The correlation between dominant hand grip and upper body strength was r=0.41 (p<0.001).

Lung function was assessed as in Study 1, only assessing forced expiratory volume per 1 second (FEV), but not forced vital capacity (FVC).

Body masculinity was calculated similar to Study 1. The variables forearm circumference, biceps circumference, chest circumference, waist-to-hip ratio, bust-to-underbust ratio, and leg-to-body ratio were extracted as automatic measurements from the body scanner (according to ISO 20685:200) and included as the average of the z-standardized values from three body scans. For forearm and biceps circumference we included the maximum value of measurements from the left and right arm. Shoulder breadth and height were measured manually. The PCA over all variables explained 43% of the total variance.

Facial masculinity and *BMI-controlled body masculinity* were assessed exactly as in Study 1.

3.1.3 Statistical Analyses

We z-standardized all variables prior to analysis and graphically inspected the data for normality and outliers. From the original sample (165 males) we excluded 1 influential case based on outlier values in body appearance (values beyond 1.5 * interquartile range above or below the third or first quartile) that showed high leverage in regression diagnostics (graphical inspection of the bivariate distribution between residuals and leverage of data points; distribution of Cook's distances; see figure S1.3 in the supplementary material S1). For the resulting sample (164 males) we computed zero-order Pearson correlation coefficients between all variables and their bootstrapped 95% confidence intervals. Again, we adjusted all p-values for false discovery rate (FDR, Benjamini & Hochberg, 1995) and additionally computed partial Pearson correlation coefficients controlling for age (ethnicity was not a varying factor in this sample). We checked whether changes in statistical significance (in terms of a threshold p<0.05) occurred between the matrices of zero-order and partial correlations and assessed their similarity via the congruency coefficient Rc (see above).

3.2 Results

Self-perceived attractiveness and self-perceived mating success were significantly correlated (r=0.47, p<0.001). Self-perceived mating success was stronger correlated to objective measures of the body of participants (body masculinity, lung function, and height) and other-rated attractiveness than self-perceived attractiveness. Notably, other people's ratings of the body were unrelated to self-perceived measures of both attractiveness and mating success (see Table 4) and other-rated facial attractiveness only showed a significant correlation with self-perceived mating success. Measures of other-rated body and facial attractiveness were however significantly correlated (r=0.40, p<0.001), which is consistent with the one ornament hypothesis (Grammer et al., 2003; Thornhill & Grammer, 1999; but see Honekopp et al., 2007) and thus indicative of the validity of the ratings.

Measures of self-perceived attractiveness were positively correlated with extraversion and inversely with neuroticism, while self-perceived mating success was correlated with conscientiousness. Extraversion was furthermore positively related to various measures of physical masculinity, while neuroticism showed a negative correlation with height (see Table 5). Measures of Dominance, Shyness, and Narcissism were largely unrelated to physical masculinity or strength, apart from a negative correlation between Shyness and physical strength and a positive correlation among Dominance and BMI-controlled body masculinity. Sociosexual Orientation, on the other hand, correlated positively with a variety of physical masculinity and strength measures, notably only the facets of Attitude and Behavior.

The similarity between partial correlations controlling for age and zero-order correlations was highly significant (Rc=0.999; 95% CI= [0.999; 1]; p<0.001). Changes of statistical significance between partial- and zero-order correlations occurred in six cases, four cases reflecting relationships between personality and formidability / physical attractiveness, namely among Sociosexual Behavior and self-perceived attractiveness, Shyness and strength, Neuroticism and residual self-perceived attractiveness (corrected for rated facial attractiveness), and Neuroticism and height. Results of

correlations based on single anthropometric measures of the body and the face of individuals can be found in the supplementary material (S4).

Table 4

Pearson correlation coefficients between self-rated and objective formidability measures

	SP attractiveness	SP mating success
Other-rated facial attractiveness	0.15 [-0.03;0.33]	**0.30 [0.18;0.42]**
Other-rated body attractiveness	0.07 [-0.10;0.22]	0.11 [-0.07;27]
Body masculinity	0.09 [-0.06;0.20]	**0.20 [0.06;0.33]**
BMI-controlled body masculinity	**0.25 [0.05;0.38]**	**0.35 [0.19;0.50]**
Upper body size	0.08 [-0.10;0.18]	**0.20 [0.07;0.31]**
Facial masculinity	-0.01 [-0.20;0.23]	0.09 [-0.06;0.24]
Strength	0.13 [-0.08;0.23]	*0.18 [0.04;0.31]*
Lung function	**0.19 [0.02,0.31]**	**0.32 [0.21;0.45]**
Height	0.09 [-0.07;0.22]	**0.25 [0.12;0.38]**
BMI	-0.05 [-0.20;0.05]	0.01 [-0.13;0.15]

Significant results in terms of FDR-adjusted p-values <.05 are displayed in bold, confidence intervals not containing the value 0 in italics. SP = Self-perceived.

Table 5

Pearson correlation coefficients between formidability and personality measures from Study 2

	Big 5					Additional Personality				Sociosexual Orientation			
	E	A	N	O	C	D	S	NA	NR	SOI-R	S-A	S-B	S-D
Self-perceived attractiveness	**0.25 [0.13; 0.40]**	0.05 [-0.07; 0.21]	**-0.22 [-0.36; -0.06]**	0.04 [-0.11; 0.16]	0.13 [-0.06; 0.25]	**0.35 [0.18; 0.47]**	-0.18 [-0.31; -0.01]	**0.33 [0.19; 0.44]**	0.00 [-0.15; 0.17]	**0.28 [0.14; 0.43]**	**0.23 [0.07; 0.39]**	**0.19 [0.05; 0.33]**	**0.20 [0.02; 0.34]**
Residual self-perceived attractiveness (corrected for rated body attractiveness)	**0.22 [0.08; 0.38]**	0.07 [-0.06; 0.23]	**-0.20 [-0.34; -0.05]**	0.05 [-0.09; 0.16]	0.16 [-0.03; 0.30]	**0.31 [0.16; 0.42]**	-0.14 [-0.27; 0.02]	**0.35 [0.22; 0.44]**	0.00 [-0.17; 0.18]	**0.25 [0.09; 0.41]**	**0.19 [0.02; 0.36]**	*0.17 [0.01; 0.30]*	**0.20 [0.05; 0.35]**
Residual self-perceived attractiveness (corrected for rated facial attractiveness)	**0.23 [0.08; 0.37]**	0.03 [-0.09; 0.17]	*-0.19 [-0.32; -0.03]*	0.08 [-0.07; 0.19]	0.13 [-0.05; 0.26]	**0.32 [0.16; 0.40]**	-0.15 [-0.28; 0.02]	**0.35 [0.20; 0.45]**	0.02 [-0.14; 0.19]	**0.25 [0.10; 0.40]**	**0.19 [0.03; 0.34]**	*0.16 [0.02; 0.30]*	**0.21 [0.05; 0.35]**
Self-perceived mating success	**0.40 [0.27; 0.55]**	0.04 [-0.12; 0.14]	**-0.29 [-0.43; -0.11]**	0.05 [-0.10; 0.18]	**0.20 [0.04; 0.34]**	**0.48 [0.34; 0.58]**	**-0.35 [-0.49; -0.20]**	**0.31 [0.14; 0.44]**	-0.02 [-0.16; 0.18]	**0.36 [0.21; 0.49]**	**0.30 [0.15; 0.43]**	**0.35 [0.19; 0.47]**	*0.17 [0.02; 0.33]*
Other-rated facial attractiveness	0.09 [-0.04; 0.23]	0.17 [0.00; 0.35]	-0.10 [-0.24; 0.05]	-0.14 [-0.28; 0.01]	0.03 [-0.11; 0.20]	0.05 [-0.08; 0.17]	-0.05 [-0.18; 0.08]	-0.05 [-0.17; 0.09]	-0.08 [-0.21; 0.07]	0.10 [-0.04; 0.24]	0.05 [-0.11; 0.15]	0.10 [-0.03; 0.24]	0.07 [-0.09; 0.23]
Other-rated body attractiveness	0.01 [-0.10; 0.17]	-0.06 [-0.22; 0.05]	0.05 [-0.06; 0.19]	-0.14 [-0.28; 0.05]	0.01 [-0.14; 0.20]	-0.04 [-0.18; 0.08]	-0.04 [-0.16; 0.09]	0.10 [-0.07; 0.24]	0.05 [-0.11; 0.18]	0.00 [-0.17; 0.15]	-0.04 [-0.18; 0.11]	-0.05 [-0.25; 0.13]	0.09 [-0.08; 0.23]
Body masculinity	**0.28 [0.14; 0.42]**	0.02 [-0.12; 0.13]	-0.10 [-0.26; 0.01]	-0.07 [-0.17; 0.06]	-0.03 [-0.18; 0.09]	*0.17 [0.04; 0.31]*	*-0.15 [-0.28; -0.05]*	0.03 [-0.11; 0.19]	0.03 [-0.14; 0.18]	**0.26 [0.12; 0.41]**	**0.24 [0.12; 0.40]**	**0.33 [0.19; 0.48]**	0.02 [-0.13; 0.18]
BMI-controlled body masculinity	**0.20 [0.04; 0.33]**	0.06 [-0.08; 0.16]	*-0.18 [-0.28; -0.02]*	-0.09 [-0.23; 0.03]	-0.07 [-0.20; 0.06]	**0.22 [0.03; 0.36]**	-0.12 [-0.26; 0.04]	0.10 [-0.04; 0.22]	0.00 [-0.12; 0.16]	**0.28 [0.15; 0.41]**	**0.29 [0.12; 0.44]**	**0.27 [0.16; 0.39]**	0.07 [-0.05; 0.24]
Upper body size	**0.25 [0.13; 0.38]**	0.02 [-0.15; 0.14]	-0.07 [-0.24; 0.08]	-0.03 [-0.14; 0.11]	-0.01 [-0.15; 0.10]	0.14 [-0.01; 0.29]	-0.11 [-0.23; 0.01]	0.03 [-0.11; 0.19]	0.01 [-0.19; 0.19]	**0.26 [0.12; 0.41]**	**0.19 [0.09; 0.33]**	**0.32 [0.18; 0.46]**	0.07 [-0.11; 0.23]

	E	A	N	O	C	D	S	NA	NR	SOI-R	S-A	S-B	S-D
Facial masculinity	0.06 [-0.10; 0.24]	-0.11 [-0.25; 0.05]	-0.07 [-0.27; 0.13]	-0.17 [-0.32; 0.00]	-0.07 [-0.21; 0.11]	0.09 [-0.11; 0.28]	-0.11 [-0.30; 0.04]	-0.10 [-0.29; 0.06]	0.03 [-0.17; 0.20]	0.09 [-0.04; 0.24]	0.07 [-0.07; 0.24]	0.13 [-0.02; 0.29]	0.01 [-0.13; 0.19]
Strength	0.12 [-0.05; 0.26]	-0.08 [-0.21; 0.08]	*0.18 [-0.33; -0.02]*	0.01 [-0.13; 0.16]	-0.02 [-0.19; 0.13]	0.13 [-0.03; 0.25]	**-0.19 [-0.37; 0.00]**	0.03 [-0.10; 0.17]	-0.06 [-0.23; 0.13]	*0.17 [0.05; 0.33]*	**0.19 [0.06; 0.33]**	**0.22 [0.08; 0.37]**	-0.03 [-0.16; 0.12]
Lung function	**0.21 [0.02; 0.32]**	0.13 [0.00; 0.26]	-0.14 [-0.27; 0.00]	-0.06 [-0.20; 0.07]	-0.03 [-0.17; 0.10]	0.14 [-0.02; 0.30]	-0.11 [-0.25; 0.04]	0.10 [-0.08; 0.27]	-0.07 [-0.21; 0.07]	0.10 [-0.08; 0.27]	0.13 [-0.07; 0.29]	0.11 [-0.04; 0.24]	-0.03 [-0.17; 0.1]
Height	0.12 [-0.03; 0.23]	0.17 [0.03; 0.29]	**-0.18 [-0.32; -0.01]**	-0.08 [-0.26; 0.09]	-0.10 [-0.23; -0.01]	0.10 [-0.06; 0.26]	-0.06 [-0.19; 0.08]	-0.08 [-0.24; 0.06]	-0.08 [-0.20; 0.04]	*0.16 [0.03; 0.26]*	**0.22 [0.08; 0.34]**	0.11 [-0.02; 0.23]	0.01 [-0.12; 0.14]
BMI	**0.21 [0.06; 0.34]**	-0.02 [-0.12; 0.11]	-0.01 [-0.15; 0.13]	-0.02 [-0.14; 0.10]	0.01 [-0.14; 0.13]	0.07 [-0.10; 0.22]	-0.10 [-0.22; 0.02]	-0.03 [-0.16; 0.13]	0.03 [-0.12; 0.18]	0.13 [-0.03; 0.30]	0.10 [-0.08; 0.29]	**0.22 [0.06; 0.37]**	-0.03 [-0.16; 0.15]

Significant results in terms of FDR-adjusted p-values <.05 are displayed in bold, confidence intervals not containing the value 0 in italics. E= Extraversion, A= Agreeableness, N= Neuroticism, O= Openness, C= Conscientiousness, D= Dominance, S= Shyness, NA= Narcissistic Admiration, NR= Narcissistic Rivalry, SOI-R= revised Sociosexual Orientation Inventory, S-A= Sociosexual Attitudes, S-B= Sociosexual Behavior, S-D= Sociosexual Desire

4. Discussion

Consistent with previous research, we found evidence for a relationship between self-rated attractiveness and various personality measures. With regard to objective indicators of formidability and other-rated physical attractiveness we mostly attained null findings, thereby not supporting the notion of reactive heritability of personality. Additional indices accounting for shared variance among self-perceived and other-rated physical attractiveness and potential confounding influences of BMI on anthropometric measurements corroborated this pattern of results. Given the limits of our sample sizes, we can, however, not rule out small effects of facultative calibration in the tested domains. Apart from the overall trend, some personality measures also did show relationships with objective trait measurements consistent with the theory of facultative calibration. This calls for a differentiated discussion of our findings. We will rest this discussion upon an integrated summary of the results of both studies given in Table 6.

Table 6

An overview of the results from Studies 1 and 2.

Women	Self-perceived attractiveness	Other-rated attractiveness	Physical strength	Anthropometric Measurements
Anger	positive association only for Success in Conflict	positive association only for Success in Conflict	no association	no association
Extraversion	positive association	no association	no association	no association
Narcissism	positive association	no association	no	no association

162

			association	
	only for the facet of Admiration			
Sociosexual orientation	positive association only for the facet of Behavior	no association	no association	no association

Men	Self-perceived attractiveness	Other-rated attractiveness	Physical strength	Anthropometric Measurements
Anger	positive association	no association	no association	positive association only for Utility of Personal Aggression
Extraversion	positive association	no association	mixed results	mixed results
Narcissism	positive association only for the facet of Admiration	no association	no association	no association
Sociosexual orientation	positive association	no association	mixed results	mixed results

With high consistency among both studies, we did hardly find any evidence for more objective, that is other-rated, physical attractiveness being a trait driving adaptive calibration of personality in men

or women (for a summary, see Figure 1). Referring to substantial correlations between self- and other-rated attractiveness, some previous studies merged these measures (Lukaszewski et al., 2014) or integrated them in a path model (Lukaszewski, 2013). Therefore we additionally assessed the effect of self-perceived attractiveness when controlling for different measures of other-rated attractiveness (i.e. residual self-perceived physical attractiveness). We found only miniscule changes in the correlations between self-rated attractiveness and all behavioral measures when controlling for other-rated attractiveness. This suggests that the relationship between physical attractiveness and personality was mainly not due to shared variance of self-perceived with more objective measures of physical attractiveness, but almost solely relied on how participants perceived themselves, regardless of how other people judged their looks. In direct comparison to Lukaszewski et al.'s (2014) result of a correlation between a composite measure of physical attractiveness and Sociosexual Orientation in men, we additionally computed composite measures of attractiveness for our samples (their z-standardized mean). Overall, the results turned out nonsignificant (supplementary material S2), although we cannot entirely rule out a small effect of composite indices based on body attractiveness. However, as we showed that rated body attractiveness was not related to Sociosexual Orientation (or had a very small effect at best), we conclude that our effect of composite physical attractiveness, and perhaps the effect of Lukaszewski (2014) as well, is foremost driven by self-perceived attractiveness. This could call for entirely different theories to explain such covariation. Haysom et al. (2015) already discussed the potential role of a positivity bias in the relationship among self-perceived attractiveness and extraversion. Also, a mediating role of self-esteem, one of the strongest correlates of self-perceived attractiveness (Feingold, 1992), could explain attractiveness-personality relationships without a mechanism of facultative calibration, as it is entailed in or correlates with personality constructs such as extraversion (Robins et al., 2001), narcissism , and sociosexuality (Jonason, Teicher, & Schmitt, 2011). At last, the correlations between self-perceived and other-rated attractiveness were also not particularly strong in our samples, especially for women in Study 1 and men in Study 2. In fact, we would not expect them to be, given a long-known meta-analytical effect size of r=0.24 for both sexes (Feingold, 1988). This could possibly be due to a flawed

operationalization of physical attractiveness by external information like anthropometric measurements or other people's judgements. An impoverishment of judgments of body attractiveness could have occurred from only rating the participants` body scans, which lack features such as skin tone or texture. As however ratings of facial attractiveness were in line with the results of body attractiveness, a disparity among self- and other-rated attractiveness could also be, as argued above, due to biases in the assessment of one's own attractiveness. These issues taken aside, effects of both, internal and external, representations of physical attractiveness or their shared variation would have been more compelling results in line with facultative calibration than a solitary effect of self-perceived attractiveness.

Consistent with the theory of facultative calibration, physical strength and physical masculinity were related to extraverted behavior in men. Interestingly, on a facet level analysis of Extraversion we found a relationship with being active and assertive, attributes that by definition require physical ability or imply a proneness to conflict. More prosocially orientated facets such as Gregariousness or Warmth, however, were unrelated to formidability. Hence and perhaps unsurprisingly, male formidability may be only related to aspects of extraverted behavior that are inherently linked to strength and body condition. This goes at hand with Mõttus' (2016) suggestion of a stronger consideration of personality trait diversity that is using facets or even items as predictor, when linking personality to specific outcomes. In Study 2 we also found a domain-level correlation between BFI Extraversion and male formidability, suggesting overall calibration of Extraversion. Interestingly, the NEO Extraversion facets Assertiveness and Activity are the only two facets (out of six) that are clearly represented in the item pool of the BFI (Soto & John, 2009). The domain-level relationship from Study 2 could thus be foremost driven by facultative calibration of these specific aspects of extraverted personality. More studies are needed to secure a detailed knowledge of a formidability-extraversion-relationship.

Both studies differed substantially in their results for Sociosexual Orientation. We found some, although not consistent, evidence for a relationship of Sociosexuality with physical masculinity and physical strength in men in Study 2. As to be expected, significant correlations appeared in the

facets of *Attitude* and *Behavior*, not in *Desire*. Contrary to that, the results of Study 1 did overall not support a relationship between formidability and Sociosexual Orientation. Thus, we remain with mixed evidence of whether markers of genetic quality (Gangestad & Simpson, 2000) may influence orientation towards and success in short-term mating for men. Again, more studies are needed to clarify the theoretical applicability of facultative calibration in this domain.

In addition to previous studies, we explored personality-formidability relationships for other traits of the Big Five personality spectrum. Whereas Openness to Experience, Agreeableness, and Conscientiousness were unrelated to physical formidability and other-rated attractiveness, we did find some, although again inconsistent, evidence for a negative relationship between physical formidability and Neuroticism. Furthermore, there was a relationship between physical strength and Shyness in men. Linking this back to the correlations between Assertiveness and formidability, this could reflect that stronger and more masculine men are less fearful of potential conflicts. These results are furthermore consistent with previous findings showing a negative relationship between handgrip strength and Neuroticism in men (Fink, Weege, Pham, & Shackelford, 2016).

More recent empirical work on the theory of facultative calibration has extended its scope to a relationship between Aggressiveness and Coalitional Strength in adolescents (Sell et al., 2016). As these authors point out, Coalitional Strength could be influenced by a variety of traits that are rather unrelated to body condition (e.g. specialized knowledge, skills, or mutual interests). However, opening the idea of facultative calibration to indicators of social relationships increases the potential of finding links where directions of causality are hard to identify (especially in cross-sectional data) and potentially even circular. Although Coalitional Strength could still be causing facultative calibration, effects of behavior on Coalitional Strength are equally likely, even more so than from behavior to body condition. In our study we purposely focused on testing the idea of behavioral calibration to phenotypic traits related to bodily and facial appearance and body condition. The absence of correlations among personality and facial attractiveness or facial masculinity in our results thereby matches a recent study testing facultative calibration of egalitarianism (Price et al., 2017), which did only find significant correlations for bodily formidability as well.

A caveat of using cross-sectional data to test facultative calibration (which applies to our studies as well) is that some phenotypic traits discussed as anchors of recalibration, e.g. physical strength, are somewhat plastic. Thus, the mechanism of recalibration could be, at least in some cases, reversed. For example, individuals with greater genetic dispositions to be extraverted, dominant, narcissistic, or aggressive might select their environments or elicit reactions from others in a way that they find themselves in competitive situations more often. In response to these experiences they might learn that a higher physical formidability would be beneficial to them and decide to work out in order to increase it. Similarly, more extraverted or narcissistic individuals might actively or reactively encounter social situations more often where higher attractiveness is more beneficial and as a response work out more, increase their grooming behavior or even become more likely to seek out cosmetic surgery. Such cases would explain an inverted causal direction, and they are indistinguishable in cross-sectional correlational studies. Indeed, Holtzman and Strube (2013) found stronger relationships between narcissism and effective adornment than natural beauty, and von Soest and colleagues (2009) showed that female patients undergoing cosmetic surgery were already more extraverted prior to surgery than females from a representative control sample, indicating another possibility of a reversed causal direction related to body attractiveness. Both examples are possible scenarios of a gene-environment-correlation, where a genetically influenced personality trait leads to the selection of and adaption to specific environments (Bleidorn et al., 2014). Another example of reverse causation could occur via positive ontogenetic feedback among personality traits such as Extraversion or Aggressiveness and self-esteem, which in turn may influence levels of self-perceived attractiveness.

As facultative calibration has not been proposed to be the one and only mechanism driving personality variation, a control of genetic confounders will gather advocates and opponents of the theory of facultative calibration most likely under the same umbrella (see also Lukaszewski & Roney, 2015, on this matter). The common method of assessing reactive heritability in genetically informed studies that have so far tested facultative calibration was to compare the heritability of the

personality trait before and after removing the shared genetic variance with another correlated phenotypic trait (Haysom et al., 2015; von Rueden et al., 2015). This is an interesting approach, since it indicates whether the genetic components of traits are independent or not. However, a substantial amount of shared genetic variation is still not a proof of facultative calibration, since such pleiotropy can still be due to various different mechanisms, including reverse causality and biological pleiotropy (Johnson et al., 2011; Solovieff et al., 2013). The approach of Lukaszewski and Roney (2011; see above) to assess the independence of effects due to calibration and a genetic polymorphism in the androgen receptor gene was thus commendable, however to be improved in complexity and sample size, since a single polymorphism is highly unlikely to reflect the genetic origin of complex trait variation (which is known to consist of a large number of very small additive genetic effects, see Munafò & Flint, 2011; Zietsch, 2016). In the animal literature, correlations between physiological and behavioral traits have already been theoretically, though not consistently empirically, differentiated into either trans-generational genetic effects (a pace-of-life syndrome; Réale et al., 2010) or ontogenetic adaptations (a state-behavior feedback loop; Sih et al., 2015; Wolf & Weissing, 2010). Empirically dissecting correlations based on such a two-fold theoretical classification may be of advantage for the human literature as well. A suited study design to further explore the origin of phenotypic covariation in a quasi-experimental way would be a longitudinal twin difference study (McGue et al., 2010), that could test the effect of differences in formidability or physical attractiveness among twins while controlling for genetic confounders and reversed causality due to ontogenetic changes in physical traits.

Acknowledgements

Data Availability

The data associated with this research are available at https://osf.io/vyw8b/.

References

Ackerman, R. A., Witt, E. A., Donnellan, M. B., Trzesniewski, K. H., Robins, R. W., & Kashy, D. A. (2011). What Does the Narcissistic Personality Inventory Really Measure? *Assessment*, *18*(1), 67–87. https://doi.org/10.1177/1073191110382845

Archer, J., & Thanzami, V. (2007). The relation between physical aggression, size and strength, among a sample of young Indian men. *Personality and Individual Differences*, *43*(3), 627–633. https://doi.org/10.1016/j.paid.2007.01.005

Asendorpf, J.B., & Wilpers, S. (1998). Personality effects on social relationships. J*ournal of Personality and Social Psychology*, *74*, 1531-1544.

Back, M. D., Küfner, A. C. P., Dufner, M., Gerlach, T. M., Rauthmann, J. F., & Denissen, J. J. A. (2013). Narcissistic admiration and rivalry: Disentangling the bright and dark sides of narcissism. *Journal of Personality and Social Psychology*, *105*(6), 1013–1037. https://doi.org/10.1037/a0034431

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

Bleidorn, W., Kandler, C., & Caspi, A. (2014). The Behavioural Genetics of Personality Development in Adulthood-Classic, Contemporary, and Future Trends: Behavioural genetics of personality development. *European Journal of Personality*, *28*(3), 244–255. https://doi.org/10.1002/per.1957

Buss, D.M. & Hawley, P.H. (Eds.) (2010). *The evolution of personality and individual differences* (pp. 243-279). New York: Oxford University Press.

Buss, D. M., & Penke, L. (2015). Evolutionary personality psychology. In M. Mikulincer, P. R. Shaver, M. L. Cooper & R. J. Larsen (Hrsg.): *APA Handbook of Personality and Social Psychology, Vol. 4: Personality Processes and Individual Differences* (S. 3-29). *Washington, DC:* American Psychological Association.

Campbell, W. K., Bonacci, A. M., Shelton, J., Exline, J. J., & Bushman, B. J. (2004). Psychological

    Entitlement: Interpersonal Consequences and Validation of a Self-Report Measure. *Journal of*

    Personality Assessment, 83(1), 29–45. https://doi.org/10.1207/s15327752jpa8301_04

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-*

    Factor Inventory (NEO-FFI) professional manual. Odessa, Fl.: Psychological Assessment.

Feingold, A. (1988). Matching for attractiveness in romantic partners and same-sex friends: A meta-

    analysis and theoretical critique. *Psychological Bulletin, 104*(2), 226–235.

    https://doi.org/10.1037/0033-2909.104.2.226

Feingold, A. (1992). Good-looking people are not what we think. *Psychological Bulletin, 111*(2), 304–

    341. https://doi.org/10.1037/0033-2909.111.2.304

Fink, B., Weege, B., Pham, M. N., & Shackelford, T. K. (2016). Handgrip strength and the Big Five

    personality factors in men and women. *Personality and Individual Differences, 88*, 175–177.

    https://doi.org/10.1016/j.paid.2015.09.013

Gabriel, M. T., Critelli, J. W., & Ee, J. S. (1994). Narcissistic Illusions in Self-Evaluations of Intelligence

    and Attractiveness. *Journal of Personality, 62*(1), 143–155. https://doi.org/10.1111/j.1467-

    6494.1994.tb00798.x

Gangestad, S. W., & Simpson, J. A. (2000). The evolution of human mating: trade-offs and strategic

    pluralism. *The Behavioral and Brain Sciences, 23*(4), 573-587-644.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H.

    G. (2006). The international personality item pool and the future of public-domain personality

    measures. *Journal of Research in Personality, 40*(1), 84–96.

    https://doi.org/10.1016/j.jrp.2005.08.007

Gough, H. G. (1956). *California Psychological Inventory.* Palo Alto, Ca.: Consulting Psychologists Press.

Grammer, K., Fink, B., Mller, A. P., & Thornhill, R. (2003). Darwinian aesthetics: sexual selection and

    the biology of beauty. *Biological Reviews, 78*(3), 385–407.

    https://doi.org/10.1017/S1464793102006085

Haysom, H. J., Mitchem, D. G., Lee, A. J., Wright, M. J., Martin, N. G., Keller, M. C., & Zietsch, B. P. (2015). A test of the facultative calibration/reactive heritability model of extraversion. *Evolution and Human Behavior*, *36*(5), 414–419. https://doi.org/10.1016/j.evolhumbehav.2015.03.002

Heymsfield, S. B., Scherzer, R., Pietrobelli, A., Lewis, C. E., & Grunfeld, C. (2009). Body mass index as a phenotypic expression of adiposity: quantitative contribution of muscularity in a population-based sample. *International Journal of Obesity*, *33*(12), 1363–1373. https://doi.org/10.1038/ijo.2009.184

Holtzman, N. S., & Strube, M. J. (2010). Narcissism and attractiveness. *Journal of Research in Personality*, *44*(1), 133–136. https://doi.org/10.1016/j.jrp.2009.10.004

Holtzman, N. S., & Strube, M. J. (2013). People With Dark Personalities Tend to Create a Physically Attractive Veneer. *Social Psychological and Personality Science*, *4*(4), 461–467. https://doi.org/10.1177/1948550612461284

Honekopp, J., Rudolph, U., Beier, L., Liebert, A., & Muller, C. (2007). Physical attractiveness of face and body as indicators of physical fitness in men. *Evolution and Human Behavior*, *28*(2), 106–111. https://doi.org/10.1016/j.evolhumbehav.2006.09.001

Isen, J. D., McGue, M. K., & Iacono, W. G. (2015). Aggressive-Antisocial Boys Develop Into Physically Strong Young Men. *Psychological Science*, *26*(4), 444–455. https://doi.org/10.1177/0956797614567718

Jacobs, I., & Scholl, W. (2005). Interpersonale Adjektivliste (IAL). *Diagnostica*, *51*(3), 145–155. https://doi.org/10.1026/0012-1924.51.3.145

Johnson, W., Penke, L., & Spinath, F. M. (2011). Heritability in the era of molecular genetics: Some thoughts for understanding genetic influences on behavioral traits. *European Journal of Personality, 25*, 254-266.

Jonason, P. K., Teicher, E. A., & Schmitt, D. P. (2011). The TIPI's validity confirmed: Associations with sociosexuality and self-esteem. *Individual Differences Research, 9*(1), 52-60.

Kinsey, A. C., Pomeroy, W. R., & Martin, C. E. (1948). *Sexual Behavior in the Human Male*. Philadelphia, PA: W.B. Saunders.

Landolt, M. A., Lalumière, M. L., & Quinsey, V. L. (1995). Sex differences in intra-sex variations in human mating tactics: An evolutionary approach. *Ethology and Sociobiology*, *16*(1), 3–23. https://doi.org/10.1016/0162-3095(94)00012-V

Lang, F. R., Lüdtke, O., & Asendorpf, J. B. (2001). Testgüte und psychometrische Äquivalenz der deutschen Version des Big Five Inventory (BFI) bei jungen, mittelalten und alten Erwachsenen. *Diagnostica*, *47*(3), 111–121. https://doi.org/10.1026//0012-1924.47.3.111

Lukaszewski, A. W. (2013). Testing an Adaptationist Theory of Trait Covariation: Relative Bargaining Power as a Common Calibrator of an Interpersonal Syndrome: Adaptationist theory of trait covariation. *European Journal of Personality*, *27*(4), 328–345. https://doi.org/10.1002/per.1908

Lukaszewski, A. W., Larson, C. M., Gildersleeve, K. A., Roney, J. R., & Haselton, M. G. (2014). Condition-dependent calibration of men's uncommitted mating orientation: evidence from multiple samples. *Evolution and Human Behavior*, *35*(4), 319–326. https://doi.org/10.1016/j.evolhumbehav.2014.03.002

Lukaszewski, A. W., & Roney, J. R. (2011). The Origins of Extraversion: Joint Effects of Facultative Calibration and Genetic Polymorphism. *Personality and Social Psychology Bulletin*, *37*(3), 409–421. https://doi.org/10.1177/0146167210397209

McCullough, M. E., Kurzban, R., & Tabak, B. A. (2013). Cognitive systems for revenge and forgiveness. *Behavioral and Brain Sciences*, *36*(1), 1–15. https://doi.org/10.1017/S0140525X11002160

McCullough, M. E., Rachal, K. C., Sandage, S. J., Worthington, E. L., Brown, S. W., & Hight, T. L. (1998). Interpersonal forgiving in close relationships: II. Theoretical elaboration and measurement. *Journal of Personality and Social Psychology*, *75*(6), 1586–1603.

McGue, M., Osler, M., & Christensen, K. (2010). Causal Inference and Observational Research: The Utility of Twins. *Perspectives on Psychological Science*, *5*(5), 546–556. https://doi.org/10.1177/1745691610383511

Morf, C. C., & Rhodewalt, F. (2001). Unraveling the Paradoxes of Narcissism: A Dynamic Self-Regulatory Processing Model. *Psychological Inquiry*, *12*(4), 177–196. https://doi.org/10.1207/S15327965PLI1204_1

Mõttus, R. (2016). Towards More Rigorous Personality Trait-Outcome Research: Personality traits and outcomes. *European Journal of Personality*, *30*(4), 292–303. https://doi.org/10.1002/per.2041

Munafò, M. R., & Flint, J. (2011). Dissecting the genetic architecture of human personality. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2011.07.007

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Parker, G. A. (1974). Assessment strategy and the evolution of fighting behaviour. *Journal of Theoretical Biology*, *47*(1), 223–243.

Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, *36*(6), 556–563. https://doi.org/10.1016/S0092-6566(02)00505-6

Penke, L., & Asendorpf, J. B. (2008). Beyond global sociosexual orientations: a more differentiated look at sociosexuality and its effects on courtship and romantic relationships. *Journal of Personality and Social Psychology*, *95*(5), 1113–1135. https://doi.org/10.1037/0022-3514.95.5.1113

Penke, L., Denissen, J. J. A., & Miller, G. F. (2007). The evolutionary genetics of personality. *European Journal of Personality*, *21*(5), 549–587. https://doi.org/10.1002/per.629

Penton-Voak, I. S., Jones, B. C., Little, A. C., Baker, S., Tiddeman, B., Burt, D. M., & Perrett, D. I. (2001). Symmetry, sexual dimorphism in facial proportions and male facial attractiveness. *Proceedings of the Royal Society B: Biological Sciences*, *268*(1476), 1617–1623. https://doi.org/10.1098/rspb.2001.1703

Petersen, M., Sell, A., Tooby, J., and Cosmides, L. (2010). Evolutionary psychology and criminal justice: A recalibrational theory of punishment and reconciliation. In H. Høgh-Olesen (Ed.): *Human Morality and Sociality: Evolutionary and Comparative Perspectives* (pp. 72-131). New York: Palgrave MacMillan.

Price, M. E., Dunn, J., Hopkins, S., & Kang, J. (2012). Anthropometric correlates of human anger. *Evolution and Human Behavior*, *33*(3), 174–181. https://doi.org/10.1016/j.evolhumbehav.2011.08.004

Price, M. E., Sheehy-Skeffington, J., Sidnaius, J., & Pound, N. (2017). Is sociopolitical egalitarianism related to bodily and facial formidability in men? *Evolution and Human Behavior*, *38*(5), 626–634. https://doi.org/10.1016/j.evolhumbehav.2017.04.001

Réale, D., Garant, D., Humphries, M. M., Bergeron, P., Careau, V., & Montiglio, P.-O. (2010). Personality and the emergence of the pace-of-life syndrome concept at the population level. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1560), 4051–4063. https://doi.org/10.1098/rstb.2010.0208

Revelle, W. (2016). psych: Procedures for Personality and Psychological Research (Version 1.6.9). Evanston, Illinois, USA: Northwestern University. Retrieved from https://CRAN.R-project.org/package=psych

Robins, R. W., Tracy, J. L., Trzesniewski, K., Potter, J., & Gosling, S. D. (2001). Personality Correlates of Self-Esteem. *Journal of Research in Personality*, *35*(4), 463–482. https://doi.org/10.1006/jrpe.2001.2324

Sell, A., Eisner, M., & Ribeaud, D. (2016). Bargaining power and adolescent aggression: the role of fighting ability, coalitional strength, and mate value. *Evolution and Human Behavior*, *37*(2), 105–116. https://doi.org/10.1016/j.evolhumbehav.2015.09.003

Sell, A., Tooby, J., & Cosmides, L. (2009). Formidability and the logic of human anger. *Proceedings of the National Academy of Sciences*, *106*(35), 15073–15078. https://doi.org/10.1073/pnas.0904312106

Sih, A., Mathot, K. J., Moirón, M., Montiglio, P.-O., Wolf, M., & Dingemanse, N. J. (2015). Animal personality and state–behaviour feedbacks: a review and guide for empiricists. *Trends in Ecology & Evolution*, *30*(1), 50–60. https://doi.org/10.1016/j.tree.2014.11.004

Smith, K. L., Cornelissen, P. L., & Tovée, M. J. (2007). Color 3D bodies and judgements of human female attractiveness. *Evolution and Human Behavior*, *28*(1), 48–54. https://doi.org/10.1016/j.evolhumbehav.2006.05.007

Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M., & Smoller, J. W. (2013). Pleiotropy in complex traits: challenges and strategies. *Nature Reviews Genetics*, *14*, 483-495.

Soto, C. J., & John, O. P. (2009). Ten facet scales for the Big Five Inventory: Convergence with NEO PI-R facets, self-peer agreement, and discriminant validity. *Journal of Research in Personality*, *43*(1), 84–90. https://doi.org/10.1016/j.jrp.2008.10.002

Stulp, G., Kordsmeyer, T., Buunk, A. P., & Verhulst, S. (2012). Increased aggression during human group contests when competitive ability is more similar. *Biology Letters*, *8*(6), 921–923. https://doi.org/10.1098/rsbl.2012.0591

Thornhill, R., & Grammer, K. (1999). The Body and Face of Woman. *Evolution and Human Behavior*, *20*(2), 105–120. https://doi.org/10.1016/S1090-5138(98)00044-0

Tiddeman, B. P., Duffy, N., & Rabey, G. (2000). Construction and visualisation of three-dimensional facial statistics. *Computer Methods and Programs in Biomedicine, 63*, 9-20.

Tooby, J., & Cosmides, L. (1990). On the universality of human nature and the uniqueness of the individual: the role of genetics and adaptation. *Journal of Personality*, *58*(1), 17–67.

Tooby, J., Cosmides, L., Sell, A., Lieberman, D., & Sznycer, D. (2008). Internal regulatory variables and the design of human motivation: A computational and evolutionary approach. In A. J. Elliot (Ed.), *Handbook of approach and avoidance motivation* (pp. 251–271). New York, NY, USA: Psychology Press.

von Rueden, C. R., Lukaszewski, A. W., & Gurven, M. (2015). Adaptive personality calibration in a human society: effects of embodied capital on prosocial traits. *Behavioral Ecology*, *26*(4), 1071–1082. https://doi.org/10.1093/beheco/arv051

von Soest, T., Kvalem, I. L., Skolleborg, K. C., & Roald, H. E. (2009). Cosmetic surgery and the relationship between appearance satisfaction and extraversion: Testing a transactional model of

personality. *Journal of Research in Personality*, *43*(6), 1017–1025.

https://doi.org/10.1016/j.jrp.2009.07.001

Wetzel, E., Leckelt, M., Gerlach, T. M., & Back, M. D. (2016). Distinguishing Subgroups of Narcissists

with Latent Class Analysis: Subgroups of narcissists. *European Journal of Personality*, *30*(4),

374–389. https://doi.org/10.1002/per.2062

Wiggins, J. S., Trapnell, P., & Phillips, N. (1988). Psychometric and Geometric Characteristics of the

Revised Interpersonal Adjective Scales (IAS-R). *Multivariate Behavioral Research*, *23*(4), 517–530.

https://doi.org/10.1207/s15327906mbr2304_8

Wolf, M., & Weissing, F. J. (2010). An explanatory framework for adaptive personality differences.

Philosophical Transactions of the Royal Society B: Biological Sciences, *365*(1560), 3959–3968.

https://doi.org/10.1098/rstb.2010.0215

Zietsch, B. P. (2016). Individual differences as the output of evolved calibration mechanisms: does the

theory make sense in view of empirical observations? *Current Opinion in Psychology*, *7*, 71–75.

https://doi.org/10.1016/j.copsyc.2015.08.014

Figure captions

Figure 1: Pearson correlation coefficient r between combined personality measures and different types of attractiveness measures for the two samples. The dashed lines indicate the threshold of statistical significance based on α=0.05 for the respective sample. If correlations were a priori hypothesized to be negative (concerning the personality variables shyness and neuroticism), they were multiplied by -1 to match the expected direction of the other correlations. The whiskers of the boxplots represent minimum or maximum values that do not exceed the range of 1.5 * interquartile range (vertical size of the box).

Figures

Figure 1