

University of Groningen

The development of animal personality

Stamps, Judy; Groothuis, Ton G.G.

Published in:
 Biological Reviews

DOI:
[10.1111/j.1469-185X.2009.00103.x](https://doi.org/10.1111/j.1469-185X.2009.00103.x)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
 Publisher's PDF, also known as Version of record

Publication date:
 2010

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Stamps, J., & Groothuis, T. G. G. (2010). The development of animal personality: Relevance, concepts and perspectives. *Biological Reviews*, 85(2), 301-325. <https://doi.org/10.1111/j.1469-185X.2009.00103.x>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

The development of animal personality: relevance, concepts and perspectives

Judy Stamps and Ton G. G. Groothuis

(Received 23 March 2009; revised 07 June 2009; 07 September 2009; accepted 12 September 2009)

ABSTRACT

Recent studies of animal personality have focused on its proximate causation and its ecological and evolutionary significance, but have mostly ignored questions about its development, although an understanding of the latter is highly relevant to these other questions. One possible reason for this neglect is confusion about many of the concepts and terms that are necessary to study the development of animal personality. Here, we provide a framework for studying personality development that focuses on the properties of animal personality, and considers how and why these properties may change over time. We specifically focus on three dimensions of personality: (1) contextual generality at a given age or time, (2) temporal consistency in behavioural traits and in relationships between traits, and (3) the effects of genes and experience on the development of personality at a given age or life stage. We advocate using a new approach, contextual reaction norms, to study the contextual generality of personality traits at the level of groups, individuals and genotypes, show how concepts and terms borrowed from the literature on personality development in humans can be used to study temporal changes in personality at the level of groups and individuals, and demonstrate how classical developmental reaction norms can provide insights into the ways that genes and experiential factors interact across ontogeny to affect the expression of personality traits. In addition, we discuss how correlations between the effects of genes and experience on personality development can arise as a function of individuals' control over their own environment, *via* niche-picking or niche-construction. Using this framework, we discuss several widely held assumptions about animal personality development that still await validation, identify neglected methodological issues, and describe a number of promising new avenues for future research.

Key words: development, behavioural syndromes, coping style, reaction norms, contextual generality, temporal consistency, gene environment correlations, gene environment interactions, niche picking, niche construction, ontogeny.

CONTENTS

I. Introduction	302
II. Basic Concepts	304
(1) Overview	304
(2) Behaviour across contexts	304
(a) Contextual generality and contextual plasticity	305
(b) Using contextual reaction norms to study variation in behaviour across contexts	306
(3) Temporal consistency and personality traits	308
(a) Stability and ontogenetic change	308
(b) A lexicon	308
(4) Criteria for animal personality traits	310
III. Gene-Environment Interactions and Correlations	311
(1) Developmental reaction norms for personality studies	311

Address for correspondence: E-mail: jastamps@ucdavis.edu

(2) Using genotypes as ‘replicate individuals’ to measure the effects of I, E and IxE on the development of personality traits	312
(3) Niche-picking and niche-construction	313
IV. Further Questions about Personality Development	314
(1) Are extreme individuals more temporally stable than intermediate individuals?	314
(2) Does artificial selection change developmental plasticity or contextual plasticity?	315
(3) Is contextual plasticity a ‘meta-personality’ trait?	316
(4) How does experience affect correlations between personality and physiological traits?	317
(5) Can we predict the ages/life stages when experiential factors are likely to affect differential consistency?	317
(6) How does spatial and temporal variation in environmental factors affect personality development?	318
(7) When during ontogeny are personality traits likely to be temporally unstable?	319
(8) How can we simultaneously analyse contextual and temporal components of personality?	319
V. Conclusions	320
VI. Acknowledgements	320
VII. References	320

I. INTRODUCTION

Over the past decade students of animal behaviour have become increasingly interested in individual differences in behaviour that are consistent across time and/or across contexts, as evidenced by the rapidly growing literature on animal personality, temperament, coping styles, and behavioural syndromes (for recent reviews see Dall, Houston & McNamara, 2004; Sih, Bell & Johnson, 2004a; Sih *et al.*, 2004b; Koolhaas *et al.*, 2007; Reale *et al.*, 2007; Sih & Bell, 2008). ‘Personality’ is a term borrowed from psychology, where it refers to underlying behavioural tendencies that differ across individuals, that are consistent within individuals over time, and that affect the behaviour that is expressed in different contexts (Caspi, Roberts & Shiner, 2005; Reale *et al.*, 2007). ‘Temperament’ and ‘coping styles’ are terms that initially developed in separate literatures, but whose meaning has recently converged with the definition for personality indicated above (Gosling, 2001; Caspi *et al.*, 2005; Koolhaas *et al.*, 2007; Reale *et al.*, 2007). By contrast, the term ‘behavioural syndrome’ is currently defined as individual differences in behaviour patterns that are either correlated across time (e.g. aggressiveness of the same set of individuals before *versus* after sexual maturity), or across contexts (e.g. activity of the same set of individuals in their home cages *versus* in an open field) (Bell, 2007; Sih & Bell, 2008). Hence, any behaviour that satisfies the criteria for personality also satisfies the criteria for a behavioural syndrome, but the reverse is not the case. Since our goal is to grapple with the complex problem of describing the development of individual differences in behavioural traits that are both contextually general and temporally consistent, here we rely on the term animal personality.

Thus far, animal personality traits have been documented in many vertebrates, including salamanders (Sih, Kats & Maurer, 2003), fish (Wilson *et al.*, 1993; Overli, Winberg & Pottinger, 2005), lizards (Cote & Clobert, 2007), birds (Groothuis & Carere, 2005), rodents (Koolhaas *et al.*, 1999)

and other mammals [e.g. mink (Malmkvist & Hansen, 2002), bighorn sheep (Reale *et al.*, 2000), see also Gosling (2001) and Sih & Bell (2008)]. Now, interest is turning to proximate and ultimate questions about personality. We suggest that an understanding of the development of personality is critical for an understanding of questions about its causation, evolution and function. First, information about the ontogenetic stages when personality traits are first expressed or change can provide valuable insights into the physiological processes that are responsible for the organisation and stability of personality traits. If, for example, major changes in personality in a species of mammal occur around the time of maturity, gonadal hormones might well be involved in its adult organisation. Since the effects of gonadal hormones on brain-behaviour relationships in vertebrates such as birds and mammals are already well known (Nelson, 2005), longitudinal studies of personality traits over time may provide important insights for researchers interested in the proximate mechanisms that generate personality in these taxa. Second, studies of the evolution of personality require information about developmental mechanisms, both genetic (Dingemanse *et al.*, 2002; Van Oers *et al.*, 2005a; Sinn, Apiolaza & Moltschanivskyj, 2006) and epigenetic (Bossdorf, Richards & Pigliucci, 2008; Champagne, 2008; Crews, 2008), that might contribute to the inheritance of personality traits. Genes, maternal and intergenerational environmental effects can all influence the effects of natural selection on any phenotypic trait, including personality traits (e.g. Moore, Wolf & Brodie, 1998). In addition, information about the effects of biologically relevant experience on correlations among personality traits can contribute to our understanding of how evolutionary and functional constraints and trade-offs affect those correlations.

Even behavioural ecologists who study the functional significance of animal personality traits rely on tacit assumptions about developmental processes. Field studies of the fitness consequences of animal personality traits are often conducted in areas in which environmental conditions vary within the

lifetimes of the study animals (Dingemanse *et al.*, 2004; Boon, Reale & Boutin, 2008). In this situation, investigators often assume that consistent individual differences in behaviour are stable throughout life (Smith & Blumstein, 2008), e.g. that individuals who were relatively bold during years when predators were present are also relatively bold during years when predators were absent (Reale & Festa-Bianchet, 2003). However, this assumption needs to be validated, given the distinct possibility that at least some of the individuals in a population might change their behaviour in response to major changes in food levels, population density, group sizes, exposure to predators or other environmental or social factors with strong effects on fitness. Changes in personality as a result of experience might very well be adaptive (e.g. Bell & Sih, 2007), but so little is known about the effects of experiential factors on correlations among personality traits that generalizations on this point are currently impossible.

Unfortunately, while an understanding of development can shed light on many important questions about personality, empirical studies on the development of personality in animals are still quite rare, with the possible exception of studies of humans and other primates. Thus far, studies of personality development fall into three broad categories: those that consider how personality changes over the course of a lifetime, those that consider how specific experiential or environmental factors at one point in life affect personality later in life, and those that focus on the effect of genes on personality development. However, only a few researchers have described how correlations among different personality traits change across ontogeny (e.g. Dingemanse *et al.*, 2002; Bell & Stamps, 2004; Carere *et al.*, 2005*b*; Johnson & Sih, 2007; Weiss, King & Hopkins, 2007; Sinn, Gosling & Moltschanivskyj, 2008), or have experimentally tested whether a specific type of experience at a given age affects correlations among personality traits later in life (e.g. Sluyter *et al.*, 1996; Benus, 1999; Carere *et al.*, 2005*a*; Bell & Sih, 2007; Frost *et al.*, 2007). Rather more attention has been paid to the effects of genes on personality, which empiricists have studied using artificial selection, quantitative genetics or analyses of the effects of specific loci on personality traits (e.g. van Oers *et al.*, 2004, 2005*a*; Fidler *et al.*, 2007; Reale *et al.*, 2007; Rogers *et al.*, 2008; see also below). Indeed, we suspect that one reason for the relative shortage of empirical studies on personality development is the widespread assumption that genes play a predominant role in the development of personality, often leading to the (erroneous) assumption that personality traits must be temporally stable across ontogeny. This emphasis on the role of genes in personality is understandable, given the surge of interest in the evolution of animal personality (e.g. Nettle, 2006; Sih & Bell, 2008; Wolf *et al.*, 2007; Wolf, van Doorn & Weissing, 2008; McNamara *et al.*, 2009). However, focusing on the effects of genes on personality neglects the fact that development is the result of a continuous interaction between genes and experiential factors throughout the lifetime, a topic we discuss extensively below.

Another possible reason for the scarcity of research on personality development in animals is confusion over the different types of contextual and temporal consistency that must be specified in order to study developmental questions. Many of the terms that have been used in animal personality studies are ambiguous: e.g. the term 'behavioural plasticity' can refer to variation in behaviour as a function of the environment in which animals were raised prior to expressing the behaviour (e.g. Sinn *et al.*, 2008) or to variation in behaviour as a function of the stimuli that surround an animal at the time it expresses behaviour (e.g. Briffa, Rundle & Fryer, 2008). Similarly, behavioural biologists have recently become interested in using reaction norms (the range of phenotypes produced by individuals with the same genotype) to study behavioural traits (Stamps, 2003; Fuller, Sarkar & Crews, 2005; van Oers *et al.*, 2005*a*; Dingemanse, 2007; Zhou *et al.*, 2008, see also Sections II and III). However, this is also a source of potential confusion, e.g. between reaction norms that focus on the effects of different rearing environments on the behavioural phenotypes expressed by individuals with different genotypes later in life (e.g. Stamps, 2003), and reaction norms that focus on the effects of current environmental conditions on the phenotypes expressed by individuals at a given age or time (e.g. Nussey, Wilson & Brommer, 2007). As a result of this lack of clarity, it is difficult for researchers to envision how developmental processes might impact the design or interpretation of their own studies, let alone identify critical questions about personality development that should be addressed in future studies.

The first goal herein is to provide a framework for studies of personality development by clarifying the concepts and terminology that are central to this topic. Hence, we initially focus on concepts that help us to understand the two key attributes of personality traits: consistency in behaviour across contexts at a given time, and consistency in behaviour across time. We then show how this framework, coupled with a traditional developmental reaction norm approach, allows us to study how effects of genes, experiential factors earlier in life and interactions and correlations among genes and experiential factors affect the behaviour that individuals express in a given context at a given time. Finally, we provide a series of novel questions about personality development that can be explored using this framework.

Although we use the concept of personality to frame this article, many of the terms and ideas outlined below can be used to study the development of other behavioural or physiological traits that differ across individuals within the same population. These include behaviour patterns that are correlated across contexts or across time, but not both (e.g. traits that satisfy the definition for behavioural syndromes but not the definition for personality), as well as a wide range of behavioural and physiological traits that have been hitherto ignored in the literatures on personality, coping styles, or behavioural syndromes.

II. BASIC CONCEPTS

(1) Overview

Personality refers to individual differences in behaviour that are consistent both across contexts and across time. As such, personality is a complicated concept, because although it requires repeated measurements of the same individuals at different times and in different contexts, the focus is actually on the behaviour of individuals relative to one another, not on the absolute levels of behaviour expressed by individuals or by groups of individuals. Indeed, as we discuss below, it is perfectly possible for personality to be stable over a period of time even if the scores of individual subjects, and/or the mean score for all of the subjects, change dramatically over that same period of time. This emphasis on the behaviour of individuals relative to others is reflected in the use of descriptive phrases such as ‘consistent individual differences’, and by the use of correlations across individuals in behaviour expressed in different contexts and at different times to measure personality traits.

To date, most empirical studies of animal personality have relied on two methods: coding of animal scores in standardized tests, and subjective rankings of personality traits by human observers (Gosling, 2001). Although both may generate comparable results (Feaver, Mendl & Bateson, 1986) here we focus on the first method, because researchers who currently use the ranking method explicitly assume that personality traits are expressed in different contexts and are stable across time, whereas standardized tests provide the data required to test these assumptions. A third method (‘naturalistic observations’ of behaviour, i.e. observations of unmanipulated animals in the field or in semi-natural settings) has occasionally been used for personality studies (e.g. McPhee & Quinn, 1998; Natoli *et al.*, 2005), but this method is of limited value for studying the development of personality because it is difficult to ensure that all of the subjects in a study are observed in the same contexts, at the same ages, and across the same inter-observation intervals.

(2) Behaviour across contexts

Studies of animal personality may begin by focusing on consistency in behaviour across contexts or by focusing on consistency in behaviour across time. The contextual and temporal axes of personality are equally important, but since some indices of temporal consistency involve behaviour measured in different contexts, we begin here with contextual consistency. Definitions of important terms are provided in Table 1, which is designed to help readers distinguish among terms that describe consistency in the behaviour expressed in different contexts at a given age and time (e.g. ‘contextual plasticity’) and terms that describe consistency in the behaviour expressed in one or more contexts at different ages and times (e.g. ‘differential consistency’). In addition, we include terms that are crucial for any study of behavioural development, because they describe how environmental or experiential factors in the past affect the behaviour that

animals express in a given context at a given time (e.g. ‘developmental plasticity’).

Herein, ‘context’ refers to all of the external stimuli surrounding an individual when it expresses a given behaviour. As such, context not only includes stimuli from the external environment (e.g. temperature, light conditions, structural features) but also stimuli from conspecifics (van Oers, Klunder & Drent, 2005b; Magnhagen, 2007; Webster, Ward & Hart, 2007), predators, potential food items, and any other organisms that can be detected by an animal when it performs a given behaviour. Note that because this definition of context includes the entire range of stimuli that impinge on individuals when they express behaviour, context as used here differs from some of the ways this term has been used in the personality, coping style and syndrome literatures in the past. For instance, context here encompasses both ‘situations’ (different ecological conditions) and ‘contexts’ (different functional behavioural categories) as described by Sih *et al.* (2004b). This allows us to avoid having to decide, *a priori*, the functional significance of the behaviour that the subjects perform in each context (see also below). In addition, this definition allows for a wide range of variation in contextual stimuli at the time of a test, from very subtle differences in contextual stimuli (e.g. comparing an individual’s responses to conspecific same-sex intruders of different sizes or ages in its home cage) to major differences in contextual stimuli (e.g. comparing an individual’s behaviour when it is exposed to cues from a predator in its home cage to its behaviour when it is placed in a large, novel area). Hence, for our purposes the sole requirement for studying consistency across contexts is that the behaviour of the same individuals is recorded in the presence of at least two different sets of external stimuli, regardless of whether those stimuli are produced by biotic (e.g. food, predators, conspecifics) or abiotic (e.g. size or structural features of the test arena, temperature, light) features of the external environment.

Similarly, we are interested in any behaviour patterns that animals perform in two or more different contexts. This allows us to consider both situations in which animals perform the same type of behaviour in different contexts, and situations in which animals perform different types of behaviour in different contexts, without having to decide whether the behaviour patterns in question are different from one another, based on assumptions about similarities in the form, motivational basis or functional significance of those behaviour patterns.

For example, imagine that we measure latency to approach a novel object in a home cage in one test, latency to enter a large, unfamiliar open area (an ‘open field’) in a second test, and then compute the correlation across individuals in the scores on both tests. Based on the form of the behaviour, we might assume that both tests measure the same behaviour (latency to approach). Based on assumptions about motivation or function, we might either decide that both tests measure the same type of behaviour (exploration), or that the two tests measure different types of behaviour (exploration in the first case, and boldness in the second). It

Table 1. Glossary of important terms¹

agent	an individual animal, or a group of individuals with the same genotype
context	all of the conditions and external stimuli that surround an individual at the time that it expresses behaviour
contextual generality	the extent to which scores for behaviour expressed in one context are correlated across individuals with scores for behaviour expressed in one or more other contexts, when behaviour in all of the contexts is measured at the same age and time
contextual plasticity	the extent to which the behaviour of a given agent varies across contexts, relative to the behaviour of other agents in those same contexts, when the behaviour in all of the contexts is measured at the same age and time
contextual reaction norm	a description of the patterns of variation in the behaviour expressed by an agent at a given age and time, as a function of the set of conditions (contexts) in which that agent expresses that behaviour
developmental plasticity	the extent to which the behaviour expressed by individuals with a given genotype in a given context at a given age and time varies as a function of the set of conditions experienced by those individuals before the behaviour was expressed
developmental reaction norm	a description of the patterns of variation in the behaviour expressed by individuals with a given genotype in a given context at a given age and time, as a function of the conditions experienced by those individuals before the behaviour was expressed.
differential consistency	the extent to which scores for behaviour in a given context at a given time are correlated across individuals with scores for the same behaviour in the same context at a later time. Also called broad-sense repeatability
individual stability	the extent to which an individual's behaviour in a given context changes over time.
mean-level consistency	the extent to which the mean value of the behaviour expressed by a group of individuals tested in a given context changes as a function of time
niche-construction	processes by which individuals create or encourage experiences or environments that affect their own subsequent development
niche-picking	processes by which individuals select or choose experiences or environments that affect their own subsequent development
rank-order consistency	a special case of differential consistency. Rank-order consistency indicates the extent to which the rank order of scores for a given behaviour measured in a given context at a given time are preserved if the same set of individuals is measured in the same context at a different time.
structural consistency	the extent to which correlations among behaviour patterns expressed in two or more contexts at one time are preserved when the same set of behaviour patterns is measured in the same set of contexts at a different time

¹these terms apply to any traits (e.g., including physiological traits) that vary across individuals and whose expression depends on conditions that surround the animal at the time the trait is expressed. However, since the current article focuses on personality, here we use behavioural traits to illustrate these concepts.

is not always obvious whether the behaviour patterns that animals express in the presence of different stimuli should be considered to be the same or different. Fortunately, as we will see below, the same framework for studying behavioural consistency across contexts can be used in both situations.

(a) Contextual generality and contextual plasticity

The term 'contextual generality', and its inverse, 'contextual specificity', refer to the extent to which scores in behaviour in one context are correlated across individuals with scores in behaviour in one or more other contexts, where high contextual generality indicates that the rank order of scores across individuals is maintained across contexts. Thus, contextual generality describes the behaviour of groups of individuals. Of course, in order to compute indices of contextual generality, it is necessary to measure the same individuals in two or more different contexts. These data can be used to compute another index, contextual plasticity, which refers to the behaviour of single individuals in different contexts. 'Contextual plasticity' refers to the extent to which

the behaviour of a given individual, relative to the scores of others in the same group, varies as a function of the context in which that individual expresses behaviour. For instance, one individual might have a relatively low latency to attack a conspecific opponent in its home cage and also have a relatively low latency to attack a conspecific in a novel arena (low contextual plasticity), whereas another individual might have a relatively low attack latency in its home cage but a relatively high attack latency in a novel arena (high contextual plasticity). There is a direct relationship between contextual generality at the group level and contextual plasticity at the individual level, because a high value of contextual generality for the group as a whole requires that most of the individuals within the group have low contextual plasticity. Thus far, virtually all studies of the effects of context on behaviour in the animal personality literature have focused on contextual generality, but as we discuss below, contextual plasticity is also worthy of attention.

One clear message from recent empirical studies of personality in animals is that intuition provides an unreliable

guide to which behavioural traits or tendencies are likely to be contextually general and which are likely to be contextually specific (for discussion of this topic, see Coleman & Wilson, 1998; van Oers *et al.*, 2005b; Bell, 2007). Traditionally, ethologists have tended to assume that behaviour patterns that appear to serve the same function are controlled by the same underlying motivational system (e.g. aggression, exploration, anti-predatory behaviour; for an influential paper, see Baerends, 1976). Under that assumption, one would expect behaviour patterns within the same functional system to be correlated with one another across individuals. Sometimes this is the case: in three-spined stickleback, *Gasterosteus aculeatus*, scores on behaviour in one ‘exploratory’ context (patterns of movement in an open field) are correlated with scores in a different context (response to a novel object in a familiar tank) (Dingemanse *et al.*, 2007). However, in the great tit, *Parus major*, a similar correlation was found in one selection line, but not in another (Carere *et al.*, 2005a). In dumpling squid, *Euprymna tasmanica*, scores on two tests that *a priori* were assumed to reflect ‘boldness’ were uncorrelated with one another in both juveniles and adults (Sinn & Moltschanivskyj, 2005; Sinn *et al.*, 2008).

Intuition has also proven untrustworthy in predicting correlations across individuals in behaviour patterns that are expressed in very different functional contexts. Indeed, the recent surge of interest in personality by animal behaviourists has, in large part, been driven by surprise at detecting correlations across individuals in behaviour expressed in different functional contexts (Bell, 2007). Examples include correlations between behaviour expressed in the presence of conspecific intruders (‘aggressiveness’) and behaviour in potentially dangerous situations (‘boldness’) (Huntingford, 1976), or between responses to potential food items (‘foraging’) and responses to potential mates (‘reproduction’) (Johnson & Sih, 2005). Because it is inadvisable to rely on assumptions about the contextual generality of behaviour, one of the first tasks in any study of animal personality is to document correlations across contexts in behaviour. The nature of these correlations may then guide us in unraveling the underlying (developmental) mechanisms that control those behaviours, and test the assumption that behaviours that seem to belong to the same functional system are generated and maintained by the same underlying physiological mechanisms.

(b) *Using contextual reaction norms to study variation in behaviour across contexts*

‘Reaction norms’ are defined as the range of phenotypes that can be produced by individuals with a given genotype, where ‘genotype’ indicates all of the genes within an individual (Schlichting & Pigliucci, 1998). Defining genotype in terms of all of the genes in an individual instead of focusing on specific genetic loci is advisable because the effects of genes on behavioural phenotypes are sensitive to variation in genetic background (other genes within the same individual) (Hofmann, 2003; Anholt, 2004), and because behavioural traits, including personality traits, are typically influenced

by many genes of small effect (Munafo *et al.*, 2003; Kendler & Greenspan, 2006; Shifman *et al.*, 2008). When reaction norm approaches are used to study individual differences in traits for the members of a single population, clones, inbred strains or lines, or (more approximately) selected lines or siblings may be used as representatives of the same genotype (Fuller, Sarkar & Crews, 2005; Kingsolver *et al.*, 2006; Williams, 2008).

Historically, reaction norms were developed to study how environmental conditions experienced early in life interact with genes to affect the development of morphological or life-history traits (Schmalhausen, 1949; Schlichting & Pigliucci, 1998; West-Eberhard, 2003; Fuller *et al.*, 2005). This question was typically addressed by raising individuals derived from different populations or species (‘genotypes’) in different environmental conditions, and then focusing on the mean values of morphological or life-history traits as a function of two factors: the population or species of origin, and the environmental conditions in which the subjects were raised (Gotthard & Nylin, 1995; Schlichting & Pigliucci, 1998). The result was a ‘developmental reaction norm’ for each of the genotypes in the study. Of particular interest were situations in which reaction norms differed across genotypes. When reaction norms could be modeled using linear equations, the reaction norm for each genotype could be specified by the slope and the intercept of their line (Pigliucci, 2001). Reaction norms with more complicated shapes were modeled using non-linear equations, which yielded parameter values reflecting the shape and/or the position of the reaction norms of the different genotypes (Gibert *et al.*, 1998; Pigliucci, 2001; Izem & Kingsolver, 2005). In turn, the reaction norm of a genotype could be used to estimate its ‘developmental plasticity’, a term which specifies the extent to which the trait values of a given genotype vary as a function of variation in the environments in which individuals in that genotype developed (Schlichting & Pigliucci, 1998; Pigliucci, 2001; West-Eberhard, 2003). Developmental reaction norms and developmental plasticity are core concepts for the study of personality development, and are discussed in detail in Section III.

As reaction norm perspectives began to be applied to behaviour, it became obvious that there is more than one way that environmental conditions can lead to different phenotypes for individuals with the same genotype. In addition to traditional developmental reaction norms, which describe how environmental conditions and experiences in the past affect trait values at the current time, current trait values can also vary as a function of contextual stimuli at the time of a test. To separate the latter from developmental reaction norms, we suggest a new term: ‘contextual reaction norm’. There was no need to consider contextual reaction norms when biologists focused on morphological traits, which change slowly, if at all, as a function of stimuli around the experimental subjects at the time of measurement. However, the same can not be said of phenotypic traits that do vary as a function of the stimuli surrounding an animal when those traits are expressed. Many physiological traits (e.g. stress

responses; Ellis, Jackson & Boyce, 2006) vary as a function of stimuli that impinge on the subjects at the time of trait expression, so the concept of contextual reaction norms is not restricted to behaviour.

Contextual reaction norms are of two types: those based on the mean scores of different individuals with the same genotype, and those based on the scores of individual animals (Sih *et al.*, 2004b; Reale *et al.*, 2007). In both cases, the scores for the same agent (genotype or individual) are measured in two or more different contexts. When studying behavioural traits, contextual reaction norms can be studied in relation to variation in factors in the physical environment [e.g. the ambient temperature during the test (Brodie & Russell, 1999)] or variation in factors in the social environment [e.g. the behaviour of other individuals with whom the agent is currently interacting (Taylor & Day, 2004; Smiseth, Wright & Kolliker, 2008)].

Contextual reaction norms are illustrated in Fig. 1 for a situation in which the behaviour of five different agents (roman numerals I - V) is measured in three different contexts (1, 2 or 3) at, as nearly as possible, the same age and time. The data could represent five different individuals, or the average scores of five groups of animals with genotypes I - V. In this example, contextual generality is reasonably high, i.e. the rank order of scores of these five agents is nearly the same in the three different contexts. However, the non-parallel lines in Fig. 1 indicate that contextual plasticity differs among the agents, e.g., across these three contexts, agent I is more contextually plastic than agent V.

Fig. 2 provides an example of contextual reaction norms for escape behaviour for representatives of nine genotypes (inbred lines and crosses) of Paradise fish, *Macropodus opercularis*

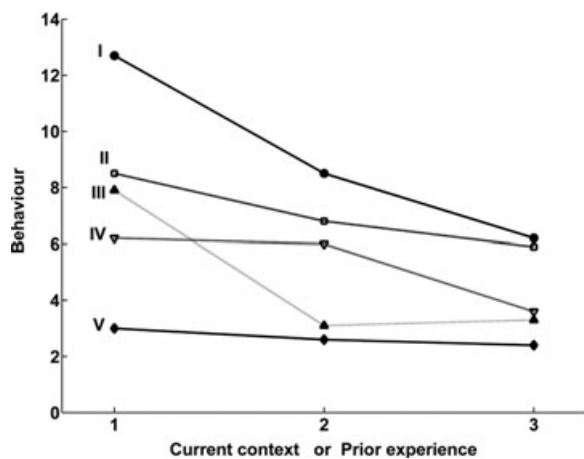


Fig. 1. A generic reaction norm diagram for individual differences in behavioural traits. For contextual reaction norms, the figure indicates the scores of five agents, I - V (either the scores of individuals or the mean scores of genotypes), each of which was tested at the same age and time in three different contexts, 1, 2 and 3. For developmental reaction norms, the diagram indicates the mean scores of individuals with five different genotypes (I - V) which were exposed to three different sets of conditions, 1, 2 or 3 before being tested in a single context.

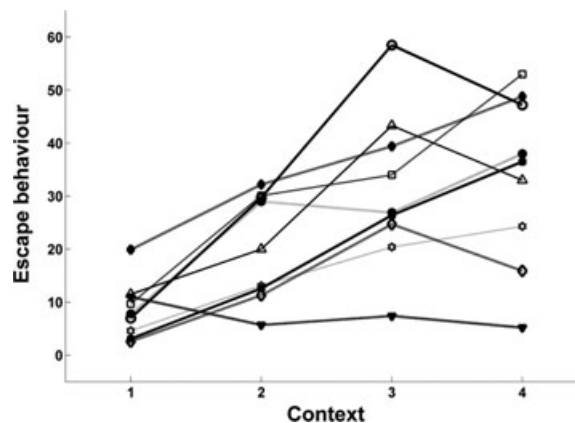


Fig. 2. Contextual norms of reaction for escape behaviour in Paradise fish, *Macropodus opercularis*. Escape behaviour (duration of rapid to-and-fro movements) for individuals from nine inbred lines and crosses between inbred lines, all of which were raised under the same conditions and then tested as adults in four different contexts: 1: home tank, 2: open field, 3: small novel tank, and 4: small novel tank + moving object. Based on data in Gerlai & Csanyi (1990).

(Gerlai & Csanyi, 1990). Contextual generality was relatively high: genotypes that had relatively low escape scores in their home tank also tended to have relatively low escape scores in the other three test contexts. At the same time, there was significant variation among the genotypes in contextual plasticity, reflected in Fig. 2 by variation among the genotypes in the shape of their contextual reaction norms. As a result, scores for escape behaviour were not only affected by genotype and by the test environment, but also by interactions between genotype and the test environment (Gerlai & Csanyi, 1990).

Contextual generality and variation among individuals in contextual plasticity can be studied using various statistical methods, including random regression models (reviewed in Nussey *et al.*, 2007). When contextual stimuli vary continuously across a gradient, it may be possible to fit random regression models using linear equations. However, more complicated models (e.g. Kirkpatrick & Heckman, 1989; Schaeffer, 2004) may be required when contexts do not fall along a continuum (e.g. as in Fig. 2), or when behaviour varies non-linearly as a function of context.

Contextual reaction norms can be useful for studying individual differences in behaviour even if contextual generality is low at the group level. In such cases, one can still ask whether individuals in the group differ with respect to their contextual plasticity, or whether low contextual generality at the group level is primarily a result of high contextual plasticity for certain categories of individuals (e.g. individuals of a given genotype, age, sex or reproductive state). Hence, contextual reaction norms can offer insights about individual differences in behaviour in different contexts regardless of the degree of contextual generality for the group as a whole.

Although contextual reaction norms can be estimated for both genotypes and individuals, it is much easier to approach the goal of “measuring behaviour in different contexts at the same age and time” using the former than using the latter. When using representatives of different genotypes to study contextual reaction norms, it is feasible to test each subject just once, and (in principle, at least) at exactly the same age and time. However, contextual reaction norm studies of individuals require a different test for each context, i.e. multiple tests conducted over non-trivial periods of time. In addition, obtaining reliable estimates of behaviour in a single context may also necessitate multiple tests of each individual (see Fleeson, 2004, for discussion of this problem when personality traits are measured in humans). Repeated testing of the same individuals means that processes such as habituation or sensitization, effects of behaviour expressed in one context on behaviour expressed in other contexts, changes in condition, nutritional state, and other state variables, and a variety of other processes can affect an individual’s scores in successive tests. In turn, inter-individual variation in temporal processes can affect estimates of personality traits at both the group and the individual level (Martin & Reale, 2008, see also below). Hence, measuring the contextual reaction norms of individuals requires that we expand our thinking into the temporal dimension.

(3) Temporal consistency and personality traits

(a) Stability and ontogenetic change

The second key criterion of personality traits is that they should be consistent within individuals for a period of time (Caspi *et al.*, 2005; Reale *et al.*, 2007). By contrast, the phrase ‘development of personality’ implies that the behaviour expressed in a given context, or correlations between behaviour patterns expressed in different contexts, may change over the course of an animal’s lifetime. Hence, anyone interested in the development of animal personality traits must consider temporal consistency over two different intervals: short intervals to determine whether behaviour is sufficiently consistent across time to be included in a study of personality, and longer intervals to determine how behaviour changes over the course of a lifetime.

The inter-test or inter-observation intervals used in studies of animal personality may be measured using absolute units of time (weeks, months, years) or using units standardized to the life history of the study species (e.g. proportion of the time from birth/hatching to sexual maturity). For comparative studies of species with similar life histories, absolute time units may be adequate, e.g. Uher, Asendorpf & Call (2008) measured personality traits of four species of long-lived great apes over the same (two week) inter-test interval. However, comparative studies of personality development involving a wider range of taxa are likely to require standardized inter-test intervals. For instance, relative to age at maturity for both species, two weeks in the life of a dumpling squid *Euprymna tasmanica* (Sinn & Moltschaniwskyj, 2005) is roughly equivalent to three years in the life of a human

(Robson & Wood, 2008)! In addition, information about age of maturity and the timing of other important life-history transitions can be used to formulate and test proximate and ultimate hypotheses about the development of personality (e.g., see Section IV.7). Hence, we suggest that studies of animal personality include information on the age at maturity for their subjects. In humans, the amount of change in personality increases as a function of the amount of time between successive tests (Schuerger, Zarrella & Hotz, 1989; Roberts & DelVecchio, 2000), and the same may be true of animals (Dingemanse *et al.*, 2002; reviewed in Bell, Hankison & Laskowski, 2009). Since behavioural tendencies that are relatively stable over short periods of time are likely to change over longer periods, researchers need to consider carefully the inter-test or inter-observation intervals that are appropriate for their species and for the questions addressed in their study.

(b) A lexicon

A major challenge for students of the development of animal personality is determining how to measure temporal changes in behavioural traits and relationships among them. Fortunately, psychologists studying personality development in humans have already confronted this challenge, and have assembled a lexicon to describe a number of different ways of measuring temporal changes in behaviour (Caspi & Roberts, 1999, 2001; Roberts, Caspi & Moffitt, 2001; De Fruyt *et al.*, 2006). For current purposes, four of these terms are most relevant: ‘mean-level consistency’ (also called normative consistency), ‘differential consistency’ (also called repeatability), ‘structural consistency’, and ‘individual stability’. The first three of these terms summarize patterns of temporal consistency in behaviour for the members of a group; by contrast, individual stability describes the temporal consistency of behaviour at the level of the individual.

Mean-level consistency is the traditional method of measuring temporal changes in behaviour. It estimates the extent to which the mean value for the score for behaviour expressed in a given context changes when the same set of individuals is tested in the same context at a later time. Mean-level consistency provides a ‘backdrop’ for any study of personality development, by indicating the general pattern of change in behaviour for all of the subjects over the inter-test interval. In animals, dramatic changes in mean-level values of animal personality traits can occur as a function of a number of factors, including age (Carere *et al.*, 2005b), life-history stage, and season (Dingemanse *et al.*, 2002), so it is important to ensure that indices used to measure other types of temporal consistency are not affected by temporal changes in mean trait values.

Differential consistency describes the extent to which subjects maintain their behaviour scores in a given context across time, relative to the scores of all of the other individuals in the same study. High levels of differential consistency can occur even if the behaviour of every individual in the sample changes over time, since this term refers to the consistency of behavioural differences among individuals,

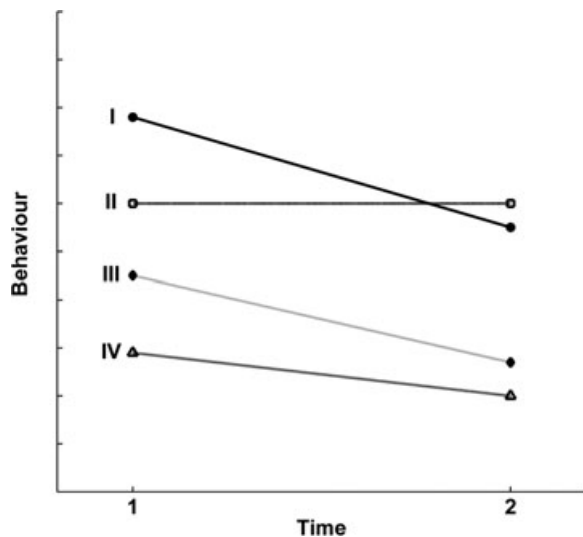


Fig. 3. Temporal changes in personality traits. Scores for behaviour expressed in a given context are indicated for four agents, I-IV (either the scores of individuals or the mean scores for individuals with the same genotype) for tests conducted at two different times (1,2). Despite a decline in the mean score from time 1 to time 2 (low mean-level consistency), differential consistency is reasonably high across this test-retest interval. At the individual level, individual stability is higher for agent II than for agent I.

not their absolute scores. High differential consistency can occur even if mean-level consistency is low (e.g. Fig. 3). Because differential consistency is computed using data collected on interval scales, this index not only indicates whether some individuals have higher or lower scores than others, but also the extent to which their scores differ from one another. Rank-order consistency is a special case of differential consistency. Because this index is based on an ordinal scale, it describes the extent to which the rank order of the scores of different individuals within the same sample is preserved over time, but ignores information about the extent to which their scores differ from one another.

Differential consistency can be measured using indices of 'repeatability' in the broad sense of that term, where repeatability is measured by computing correlations across individuals for their scores on the same behaviour in the same context at different times. In the animal behaviour literature, the two most commonly used indices of repeatability are the product-moment correlation r , and the intra-class correlation τ , where τ is 'repeatability' in the narrow sense of the term, based on equations derived and used by quantitative geneticists (Lessells & Boag, 1987; Falconer & Mackay, 1996). However, since τ is very sensitive to changes in the mean value of traits in repeated samples (Hayes & Jenkins, 1997), r provides a more useful index of differential consistency than τ for inter-test intervals when mean-level consistency is low.

Structural consistency describes temporal changes in relationships among behaviour patterns that individuals express in more than one context. This term refers to

the extent to which correlations among behaviour patterns expressed in two or more contexts at one time are preserved when the same set of behaviour patterns is measured in the same set of contexts at a later time. Hence, structural consistency applies both to situations in which the same type of behaviour is expressed in different contexts and to situations in which different behaviour patterns are expressed in different contexts. This index of temporal change is crucial for studies of animal personality, because the term personality implies that if scores for behaviour expressed in different contexts are correlated across individuals at one time, those same scores will also be correlated with one another at a later time.

Note that structural consistency can vary independently of differential consistency. For instance, in juvenile brown rockfish, *Sebastes auriculatus*, feeding behaviour in the presence of a predator and feeding behaviour in the absence of a predator were significantly correlated with one another across individuals in two assays conducted 10 days apart (high structural consistency), but the scores of individual fish were not correlated with one another across the same test-retest interval (low differential consistency) (Lee & Berejikian, 2008). Brown rockfish mature at three to six years of age (Love & Johnson, 1998), so that one would expect personality traits to be both differentially and structurally consistent across a inter-test interval of only 10 days. High structural consistency coupled with low differential consistency across longer inter-test intervals tell us how personalities change over the course of a lifetime. For instance, in a population of three-spined stickleback, 'boldness' (response to food in the presence of cues from predators) and 'aggressiveness' (response to a conspecific) were positively correlated with one another across individuals in both juveniles and adults, but individuals who were highly bold and aggressive as juveniles were not necessarily also highly bold and aggressive as adults (Bell & Stamps, 2004). In the simplest situation (behaviour measured in two different contexts at two different times), structural consistency can be evaluated by comparing the strength of the correlations between the two scores in the first pair of tests with the strength of the correlations between the same two scores in the second pair of tests.

The fourth index of temporal change in personality traits, individual stability, is the only term on our list that describes the behaviour of individuals, as opposed to summarizing temporal changes in behaviour for the group as a whole. Individual stability refers to the extent to which an individual's score on a given behavioural trait, measured in a specific context at one time, changes if the same behaviour is measured in the same context at a later time (Asendorpf, 1990, 1992; Roberts, *et al.*, 2001). Regardless of the patterns of temporal consistency across the entire sample, different individuals within that sample may differ with respect to the degree to which their scores change over time. For instance, in Fig. 3 there is a general tendency for mean scores to decline from time 1 to time 2, but the score for individual I changed more over this interval than did that of individual II. Although animal behaviourists are usually not interested in

the behaviour of single subjects, indices of individual stability can be used to compare the extent of temporal change in behaviour for different subgroups within an original group or sample, e.g. males *versus* females, or individuals with extreme scores *versus* those with intermediate scores.

Several methods have been used to measure individual stability, including the individual stability statistic (ISS) (Asendorpf, 1990, 1992) and the reliable change index (RCI) (Roberts, *et al.*, 2001). Of these, the ISS is more likely to be useful for animal studies than the RCI, because the latter is based on assumptions that may not be valid for animal personality traits, e.g. that individual scores are normally distributed and that measurement error accounts for a large proportion of the variance in the scores of the same individuals at different times (Jacobson & Traux, 1991). Sinn *et al.* (2008) provide a helpful introduction to the use of ISS in animal personality studies.

Indices of individual stability estimate the amount of temporal change in an individual's behaviour in a given context over time, regardless of the direction of that change. Complications can arise, however, if the goal of a study is to describe the direction of temporal change for subsets of individuals who had extreme scores (either extremely high or extremely low) on an initial test. The reason is a statistical phenomenon called 'regression to the mean', which occurs when an individual's score on a given test is partly due to measurement error, to uncontrolled (random) events, or to its internal state, and when those errors, events or states are unlikely to be the same when the same individual is tested again (Bland & Altman, 1994). In this situation, individuals who had very high scores in an initial test would be expected, simply by chance, to have lower scores on subsequent tests, while individuals who had very low scores in an initial test would be expected by chance to have higher scores on subsequent tests. By contrast, individuals with average scores on an initial test would, by chance, be equally likely to have higher or lower scores on subsequent tests. Statisticians have developed several methods to correct for biases due to regression to the mean (reviewed in Tu & Gilthorpe, 2007), but these methods are based on underlying assumptions about the factors that are responsible for temporal changes in test scores, and it is currently unclear whether and how these assumptions apply to personality and other behavioural traits.

Finally, we can begin to appreciate how variation among individuals in the stability of their behaviour over time (individual stability) might affect the contextual reaction norms and contextual plasticity of those individuals. As was noted above, in order to measure the contextual reaction norms of individual animals, one is forced to measure their behaviour in different contexts at different times. If an individual's behaviour over the entire testing period is highly stable in each context (e.g. individual II in Fig. 3) then it would be reasonable to use its scores in different contexts to measure its contextual reaction norm and contextual plasticity. However, if an individual's behaviour over the testing period is temporally unstable in one or more of the

contexts (e.g. individual I in Fig. 3), then data collected in tests conducted in different contexts at different times would not provide an accurate estimate of its contextual reaction norm at any given time. For instance, individuals with low individual stability over the testing period might (erroneously) appear to be highly contextually plastic. This problem can not be solved by measuring differential consistency at the group level, since individual stability can vary among the individuals within a sample, even if differential consistency at the group level is relatively high. In fact, narrow-sense repeatability values (one index of differential consistency, see above) for personality traits are generally in the neighborhood of 0.30 - 0.50 (Reale *et al.*, 2007), implying that the behaviour of many individuals was temporally unstable over the inter-test intervals used in these studies.

Hence, researchers interested in studying contextual generality and contextual plasticity can not simply assume that the behaviour of every individual in every context is temporally stable across the testing period. If this assumption is invalid, researchers run the risk of conflating individual differences in contextual plasticity (variation in the behaviour expressed in different contexts at the same time) with individual differences in individual stability (variation in the behaviour expressed in a given context at different times); for further discussion of this issue, see Section IV.8. As was noted earlier, the potential effects of temporal changes in behaviour on estimates of contextual reaction norms and contextual plasticity can be circumvented by using individuals with different genotypes to study contextual reaction norms, since in this case it actually is possible to test the representatives of each genotype at the same time in different contexts by using different individuals with the same genotype.

(4) Criteria for animal personality traits

It should by now be apparent that the concept of animal personality is more complicated than intuition might suggest. The term 'personality' implies that individuals within the same group, sample or population express different levels of the same behaviour pattern, or different behaviour patterns, in the same context at a given age and time. In addition, personality assumes some degree of contextual generality: scores for the same behaviour expressed in two or more contexts, or scores for different behaviour patterns expressed in two or more contexts, should be correlated with one another across individuals or genotypes at a given age and time. Contextual generality is implicitly assumed even in studies in which a particular 'personality trait' is measured in just one context. For instance, when researchers estimate aggressiveness using a mirror test, they assume that their subjects' scores on this test would be correlated with the same individuals' responses to conspecific opponents under natural conditions, even if they do not test this assumption (e.g. Boon, Reale & Boutin, 2007). Third, personality traits are assumed to be both differentially and structurally consistent across time. That is, personality not only requires that individual scores in a given context at one time be positively correlated with their scores in the same context at a later time

(differential consistency), but also that correlations between the same behaviour in different contexts, or between different behaviours in different contexts, measured at one time will be maintained if the same behaviours in the same individuals are measured in the same contexts at a later time (structural consistency). Both types of temporal consistency need to be documented, since structural consistency can occur without differential consistency. Note that all of these criteria for personality (individual differences in behaviour, contextual generality, differential consistency and structural consistency) all apply at the group level. Regardless of whether or not behavioural traits satisfy all of these criteria, individual-level indices of behaviour (contextual reaction norms, individual stability) can still be useful for studying how different individuals or genotypes behave in different contexts at the same time, or how the behaviour expressed by individuals or genotypes in a given context changes over time.

These criteria for personality raise questions about how strong correlations across individuals, contexts and time must be for behaviour patterns in animals to be considered personality traits or behavioural syndromes (Sih & Bell, 2008). As a practical matter, given the small sample sizes, high intra-individual variability, and low statistical power that are typical of animal personality studies, correlations among behaviour patterns across contexts or across time must be at least in the moderate range (e.g. above 0.3) to reach statistical significance in most empirical studies. However, even if animal researchers had access to the large numbers of subjects available to psychologists who study human personality, it would not be advisable to specify minimum effect sizes as cut-off criteria for animal personality traits. For instance, if scores on the same test given one week apart were related across individuals with a Pearson r correlation coefficient of 0.20, should we consider this behaviour to be differentially consistent because the scores were significantly correlated with one another with a P value of 0.001, or should we consider this behaviour to be differentially inconsistent because the correlation coefficient was relatively low?

We suggest that instead of trying to devise criteria to divide personality traits from other types of behavioural traits that are more variable across contexts or across time, it might be more profitable to use the concepts required to study personality development to address general questions about the contextual generality and temporal consistency of behavioural traits. For instance, we can use the framework outlined above to ask why some behavioural traits satisfy the criteria for behavioural syndromes but not those for personality [e.g. why 'boldness' in dumpling squid is consistent across time but not across contexts (Sinn *et al.*, 2008)], whereas other behavioural traits satisfy both sets of criteria [e.g. why 'exploratory behaviour in wild great tits is correlated across both time and contexts (Verbeek, Drent & Wiepkema, 1994)]. Hence, at either the proximate or the ultimate level, we can use this framework to ask why certain behaviour patterns are more contextually general, differentially consistent or structurally consistent than others

across ontogeny, why some traits are more strongly correlated with one another across either contexts or time in one population or species than another, or when and why contextual or temporal correlations between behavioural traits change over the course of a lifetime.

III. GENE-ENVIRONMENT INTERACTIONS AND CORRELATIONS

(1) Developmental reaction norms for personality studies

Developmental reaction norms provide an ideal framework for studying one of the most challenging questions about development, namely how specific types of experiences or exposure to environmental conditions in the past affect the expression of phenotypic traits at a given age or time. When used to study the development of individual differences in behavioural traits, a developmental reaction norm study involves exposing individuals with different genotypes to different sets of conditions, after which the behaviour of all of the subjects is tested at the same age and time in the same context. Fig. 1, previously used to illustrate contextual reaction norms, can, with a few minor modifications, also serve to illustrate developmental reaction norms. In this case, scores for the behaviour expressed in a single context are measured for individuals with five different genotypes (roman numerals I - V) which have been exposed to three different sets of conditions (1, 2 or 3) prior to the test. Prior exposure to condition 1 leads to higher average scores than does exposure to condition 3. However, the five genotypes have different developmental reaction norms (non-parallel lines), indicating that genotype and experiential factors prior to the test interact to affect an individual's scores on the test. Fig. 1 also indicates that for these three developmental conditions, genotype I is more developmentally plastic than genotype V. As we will see later (Section III.3), developmental reaction norms can also be used to illustrate gene-environment correlations (rGE), i.e. correlations between an individual's genotype and its developmental environment.

Note that parallel slopes in developmental reaction norm experiments should not be taken as evidence that gene-environment interactions do not affect the development of that phenotypic trait in that species. This is because it is often difficult to predict *a priori* which experiential factors are likely to affect the development of personality traits, or when during ontogeny biologically salient experiential factors are likely to affect behavioural development (*cf.* the concept of 'sensitive periods': Bateson, 1979). As a result, lack of evidence of individual differences in developmental plasticity may indicate that the investigator tested the wrong experiential factor, or exposed the subjects to the right factor at the wrong time or for the wrong duration. In addition, as is discussed in the next section, interactions between genes and experiential factors can have strong impacts on behavioural development well before experimental subjects

are old enough to be placed in the different ‘environments’ that are of specific interest in a developmental reaction norm study. Hence, findings indicating variation among genotypes in developmental plasticity may be more informative than results that fail to detect variation across genotypes in their developmental reaction norms.

(2) Using genotypes as ‘replicate individuals’ to measure the effects of I, E and Ix E on the development of personality traits

As was noted in Section II, developmental reaction norms were originally designed to study how genes and experiential factors interact to affect the development of phenotypic traits. In the ‘ideal’ reaction norm study, a large number of individuals with different genotypes would be raised from conception under two or more sets of environmental conditions, and then their phenotypes would be measured. The results could then be used to address questions at the group level, e.g. to estimate the effects of genes (G), environmental factors (E), and the interactions between genes and environmental factors (GxE) on the development of phenotypic traits. In addition, this sort of study could be used to address questions at the genotype level, e.g. whether certain genotypes were more developmentally plastic than others.

Unfortunately, it is virtually impossible to design a developmental reaction norm experiment such that differences in the phenotypes of genotypes raised in the same environment can be directly attributed to genetic differences among those genotypes. The basic problem is that phenotypic traits are the product of continuous interactions between genes and experiential factors (including maternal effects) that begin at conception (Jablonka, 2007), but empiricists must wait until their subjects have grown to a more advanced stage of development before providing the different ‘environments’ or ‘experiences’ that are the foci of the reaction norm study.

For instance, parents with different genotypes not only produce offspring with different genotypes, but they may also differ with respect to experiential factors and environmental conditions that they provide to their offspring, e.g. substances mothers deposit in their eggs, conditions *in utero*, or variation in parental behaviour early in life (Mendl & Paul, 1991; Groothuis *et al.*, 2005; Arnold *et al.*, 2007; Curley *et al.*, 2008). Recently it has become apparent that parental effects have profound impacts on the development of behavioural traits, including personality traits, in animals (Groothuis *et al.*, 2005; Groothuis & Carere, 2005; Uller, 2008).

Since egg quality and, in many species parental care, is under the control of mothers, thus far most authors have focused on the ways that maternal effects might affect the development of personality traits (review in Reale & Dingemans, in press). For example, it has been suggested that differences in maternal androgen deposition in the eggs of lines of great tits artificially selected for bold and shy behaviour may contribute to the differentiation in personality in this species (Groothuis & Carere, 2005; Groothuis *et al.*, 2008). Some maternal effects on offspring development may

be circumvented by experimental designs that use offspring with the same father but different mothers as representatives of different genotypes (e.g. Sinn *et al.*, 2006; Zhou *et al.*, 2008). However, this practice does not control for differential maternal allocation, a process in which maternal effects from mothers vary as a function of phenotypic traits in the males with whom those females mated (Burley, 1981; Gil *et al.*, 1999; Stamps, 2003), or for situations in which experiences of fathers affect the expression of genes that they transmit to their offspring (paternal epigenetic inheritance: Crews *et al.*, 2007). Thus far, most of the research on paternal epigenetic inheritance has focused on the adverse effects of a male’s exposure to toxins on the health of his offspring (Cordier, 2008), but there are indications that paternal experiences may also affect the expression of genes that affect the behaviour of their offspring (Crews *et al.*, 2007). In that situation, of course, differences in the behaviour of offspring sired by different fathers could be affected by both experiential and genetic differences among those sires. In addition, siblings can affect one another’s development prior to birth or hatching (Clark & Galef, 1995), and siblings or other conspecifics can mutually affect one another’s behavioural development if they are group-housed before they are placed in the developmental environments of interest to the experimenter. For instance, in laboratory strains of house mice, *Mus musculus*, the genotypes of the littermates with whom a focal individual is raised have substantial effects on the expression of aggressive and reproductive behaviour of that focal individual later in life (Crews, 2008). Many of the effects of parental or personal experience on behavioural development described above were unsuspected until very recently, suggesting that other types of experiential factors with profound effects on behavioural development have yet to be discovered.

Many of the early experiential factors that affect behavioural development systematically vary across the lines, strains, clones or sibships that might be used to represent different genotypes in reaction norm studies, leading to gene-environment correlations. As a result, even if researchers were able to place their experimental subjects in their respective developmental environments immediately after birth or hatching, the phenotypes of those subjects would already have been shaped by experiential factors that were similar for individuals with the same genotype, but different for individuals with different genotypes. Hence, while a reaction norm experiment can provide valuable insights into the ways that environmental factors interact with genes to affect behaviour, it would be a mistake to assume that any differences in the behaviour of individuals with different genotypes can be attributed to differences in their genetic makeup.

Instead, we suggest that when animals with different genotypes are used in developmental reaction norm studies, they might be viewed as approximations of ‘replicate individuals’, i.e. individuals who not only have similar genotypes, but who have also experienced similar maternal effects, paternal effects and sibling effects before being placed in the different developmental conditions that are of interest

in that study. This approach relies on the assumption that variation in conditions that affect behavioural development, as well as variation across individuals in their responses to those conditions, is likely to be smaller within genotypes than across genotypes. In addition, researchers can control experimentally or statistically for some of the factors (e.g. birth or hatch order) that are known to contribute to variation in behavioural phenotypes within genotypes.

When individuals with the same genotypes are viewed as replicate individuals, developmental reaction norm studies can be used to obtain rough estimates of the effects of I (the phenotype of individuals who share the same genotype and who experienced similar conditions prior to the beginning of the study), E (the environmental condition in which individuals experienced prior to testing) and I×E (interactions between I and E on the development of the phenotypic traits of interest) on the development of personality and other traits. Because the goal has shifted from trying to identify the effects of G and G×E on phenotypic development to studying the effects of I and I×E on phenotypic development, it is also no longer critical to begin studies with animals who are as young as possible when they are first placed in different developmental environments. Instead, researchers can begin with individuals at older ages and life stages, as long as care is taken to ensure that all of the experimental subjects have been kept in conditions as similar as possible before they are placed in the different developmental environments of interest.

This is not to say that one cannot investigate the effects of genes on the development of personality, as long as the limitations of the available methods are understood. For instance, heritability estimates based on comparisons of the behaviour of genetically related individuals (e.g. parents with their offspring, or siblings with one another) can be used to study the effects of genes on personality development (e.g. Sinn, *et al.*, 2006; Saetre *et al.*, 2006), but these methods suffer from many of the same problems outlined above. Artificial selection for lines with extremely high or low values of personality traits has also been used to estimate the effects of genes on those traits (e.g. Drent, van Oers & van Noordwijk, 2003; van Oers *et al.*, 2005a, see also below). However, since selection takes place at the level of the phenotype, estimation of gene-environment correlations and interactions based on selected lines is not always easy. The use of modern genetic manipulations such as transgenic animals may be helpful, although evidence that most behavioural traits are affected by many genes of small effect (see above) suggests that it may be difficult to find single genes with large effects on personality development. However, there are hints that single genes may affect personality traits in some species. For example, insertion of the gene for the arginine vasopressin receptor from prairie voles (*Microtus ochrogaster*) into laboratory house mice induced prairie vole-like anxiety, affiliative and mating behaviours in the transgenic mice (Pitkow *et al.*, 2001).

A final, general limitation of many of the methods used to estimate the effects of genes on behaviour is that the experimental subjects are typically exposed to just one set

of conditions prior to testing. As a result, estimates of the effects of genes on behavioural phenotypes are only valid for individuals exposed to that same set of conditions (Lewontin, 1974; Stamps, 2003; West-Eberhard, 2003). However, one of the major insights to be gained from a reaction norm perspective is that the proportion of phenotypic variation that can be attributed to genetic variation is likely to vary as a function of the conditions experienced by the study animals prior to the time or age of testing. In addition to the effects of experience on behavioural development outlined above, animals with different genotypes may also systematically differ with respect to the physical or social environments in which they develop, as a result of processes such as niche-picking or niche-construction (see Section III.3). As a result, experiments in which all of the subjects are raised under a single set of environmental conditions in the laboratory may not provide accurate estimates of the effects of genes on personality traits in nature. Although tempting, finding evidence for effects of genes does not tell us much about environmental influences and *vice versa*. Moreover, both may have similar effects on personality traits, as is the case for the effects of artificial selection *versus* manipulation of the litter sex ratio on the aggressiveness of male house mice (Mendl & Paul, 1991).

(3) Niche-picking and niche-construction

Developmental reaction norms also provide a useful framework for thinking about the effects of an individual's own behaviour on its developmental environment. Correlations between an individual's genotype and its own developmental environment (gene-environment correlations, r_{GE}), or correlations between an individual's phenotype at a given age, time or life stage and its subsequent developmental environment (r_{IE}) do not arise in traditional reaction norm studies, in which subjects with particular genotypes are forced to develop in environments selected by the investigator. By contrast, the subjects of human personality studies have considerable control over their own developmental environments. Psychologists studying the development of personality have long appreciated that correlations between an individual's phenotype at a given age and the environments in which that individual subsequently develops can occur for a variety of reasons (Plomin, Defries & Loehlin, 1977). For current purposes, two processes are particularly relevant: 'niche-picking' and 'niche-construction' (also known as niche-building) (Scarr & McCartney, 1983; Rutter & Silberg, 2002; Caspi *et al.*, 2005; Jang, 2007). Niche-picking occurs when individuals with a given phenotype choose a particular set of conditions in which to live, while niche-construction occurs when individuals with a given phenotype shape the conditions in which they live.

When applied to personality development, niche-construction refers to situations in which different individuals in the same population create or encourage different sets of experiences for themselves (Plomin *et al.*, 1977). For instance, niche-construction in the social realm would occur when aggressive acts and threat displays by a focal

individual encourage conspecifics to avoid or behave submissively to that focal individual in the future (Drummond, 2006; Kutsukake & Clutton-Brock, 2008). However, niche-construction does not require overt behaviour on the part of an actor. A possible example of niche construction in which cues emitted by one individual affect the social behaviour of others is implied by studies showing that cues emitted by sick animals cause other members of the same species to avoid contact with the infected animal (Renault, Gheusi & Aubert, 2008; Richard, Aubert & Grozinger, 2008).

Niche-picking and niche-construction may have long-term effects on the developmental environment experienced by a given individual. For instance, variation in 'sociality' at the stage of life when an individual chooses its own social environment would have a major impact on that individual's subsequent social environment if more-social individuals choose to settle in large social groups or densely populated neighbourhoods, whereas less socially-inclined individuals choose to settle by themselves, in smaller groups, or at lower population densities. To our knowledge correlations between scores for sociality prior to dispersal and choice of a post-dispersal habitat have not yet been demonstrated. However, Cote & Clobert (2007) asked a related question, namely whether scores for 'social tolerance' prior to dispersal were related to the probability of dispersal from the site of origin. They found that juvenile lizards (*Lacerta vivipara*) who dispersed away from low density sites had scored higher on a test of social tolerance soon after birth than juveniles who did not disperse from low density sites, and conversely, that juveniles who dispersed away from high density sites tended to have lower scores for social tolerance than those who remained at those sites.

More generally, an important consequence of niche-picking or niche-construction is that individuals who have different phenotypes at the ages or life stages in which they begin to control their own environments will not be randomly distributed across all of the environmental and experiential factors (social and otherwise) that affect the members of their population. Instead, one would observe rIE , a correlation between individuals' phenotype at a given age or ontogenetic stage, and the type of environment in which it subsequently develops.

Niche-picking and niche-construction have important implications for the design and interpretation of laboratory studies of personality development. In particular, if niche-picking and niche-construction lead to rIE under natural conditions, then they may generate patterns of personality development that differ from those produced under the restrictive conditions of the laboratory. For instance, with reference to Fig. 1, imagine that in nature individuals with genotype III consistently choose or create environment 1 whereas individuals with genotype I consistently choose or create environment 3. In that case, individuals with genotype III would have higher scores for the behaviour of interest than individuals with genotype I. By contrast, if individuals with both genotypes were maintained under the same conditions (either 1 or 3) for extended periods in the laboratory, then

genotype III would have lower scores than genotype I in either of the two conditions.

Niche-picking and niche-construction may also help explain situations in which differential consistency declines when individuals are brought into the laboratory and maintained there for extended periods (e.g. Wilson *et al.*, 1993). For instance, if in nature individuals with genotype III choose or create environment 2 while individuals with genotype IV chose or create environment 1, then upon first arriving in the laboratory, we would expect individuals with genotype IV to have higher scores than individuals with genotype III (Fig. 1). However, if individuals with both genotypes were confined in the laboratory for an extended period in the same environment 3 and then re-tested, there would be no discernable difference in their scores. Hence, rearing individuals under 'standard' conditions in the laboratory may yield different results than when animals are allowed access to heterogeneous environments (social or physical) in either the laboratory or the field, a point developed further in Section IV.6.

IV. FURTHER QUESTIONS ABOUT PERSONALITY DEVELOPMENT

Students of personality development in animals have begun to describe differential consistency and structural consistency at different ages over the course of the lifetime, experimentally test whether specific types of experience at a given age affect personality traits later in life, and estimate the effects of genes on personality. While all of these approaches are worthwhile, there are many other important questions about animal personality traits that can be addressed using the framework outlined in this article. Examples of these sorts of questions are outlined below.

(1) Are extreme individuals more temporally stable than intermediate individuals?

Studies of personality development in children indicate that individuals who either had very high or very low scores on an initial test of 'behavioural inhibition' (a temperament trait related to shyness; see Wilson *et al.*, 1994; Fox *et al.*, 2005) had higher levels of individual stability on that trait than individuals who initially had intermediate scores on that trait (Kagan, Reznick & Snidman, 1988; Kerr *et al.*, 1994; Sanson *et al.*, 1996). In addition, Sanson *et al.* (1996) found that children with extreme scores on a trait related to cooperation were more stable over time with respect to their scores on this trait than children with intermediate scores on this trait.

Why might extreme individuals be more temporally stable than intermediates? One possibility, suggested for humans, is that extremes are more likely than intermediates to engage in niche-picking or niche-construction, and remain in environments that encourage the maintenance of their extreme phenotypes. By contrast, intermediates might be

more likely to switch from one type of environment to another over the course of their lives, and then adjust their phenotype to suit their new environment (Buss & Plomin, 1984). For instance, highly social individuals might maintain themselves in environments which afford continuous opportunities for social interactions with a wide range of conspecifics, while highly non-social individuals might maintain themselves in environments which offer low social encounter rates and extended periods of solitude. By contrast, individuals with intermediate social phenotypes might be more likely to switch social environments during their lifetimes, and then adjust their social behaviour as a result of this environmental change. Hence, this hypothesis suggests that as a result of niche-picking or niche-construction, *rIE* will be stronger for individuals with extreme scores on initial personality tests than for individuals with intermediate scores on those same tests. In that case, even if developmental plasticity were the same for every individual in the population, individuals who had extreme scores on an initial test would have more stable personalities (higher individual stability) than individuals who initially had intermediate scores on the same test.

Measuring differences in individual stability for individuals with extreme *versus* intermediate scores is complicated by the fact that many of the tests used to measure personality traits in animals are subject to ‘ceiling’ and ‘floor’ effects. For instance, it is impossible for an individual who had a latency score of 0.0 on an initial test to have a lower score on a subsequent test, whereas an individual with an intermediate score on an initial test could have either a higher or a lower score on a subsequent test. Ceiling and floor effects can be reduced by modifying experimental protocols to ensure that every individual in a sample is able to increase or decrease their scores on subsequent tests. For instance, if a highly salient stimulus elicits immediate responses (latency scores of 0.0) for the most-responsive individuals in the population, then switching to a less evocative stimulus might encourage longer latencies for every subject, and thus provide scope for the most-responsive individuals to reduce their response latencies in subsequent tests. By designing protocols that allow for scores substantially higher and lower than those observed for any of the subjects in pilot studies, it should be possible to ask whether individual stability in animals varies as a function of their initial scores on behavioural tests, and if so, whether this pattern only occurs in heterogeneous environments (where niche-picking and niche-construction are possible) or whether it also occurs in the standardized, uniform environments that are typically used for laboratory studies of personality development.

(2) Does artificial selection change developmental plasticity or contextual plasticity?

Lines produced by strong artificial selection for extreme values of personality traits have been used to study the proximate bases of personality in a wide range of animals, including house mice (e.g. Koolhaas *et al.*, 1999), chickens, *Gallus gallus* (van Hierden *et al.*, 2002), great tits (Groothuis *et al.*, 2008), Japanese quail, *Coturnix japonica* (Gil & Faure,

2007), and rainbow trout, *Oncorhynchus mykiss* (Overli *et al.*, 2005). These and other selected lines have made an number of important contributions to our understanding of animal personality. However, there is an important distinction between using selected lines to suggest hypotheses about animal personality, and using selected lines to test those hypotheses. In particular, using selected lines to test hypotheses about personality development relies on the implicit assumption that the mechanisms that are responsible for the differences in the scores of the selected lines are the same as the mechanisms responsible for the consistent individual differences in trait values in the population that gave rise to those selected lines. This assumption need not always be valid. For instance, under the view that strong directional selection for a given trait selects for changes in regulatory processes that increase the chances that individuals will express that trait (e.g. see West-Eberhard, 2003, p. 149), artificial selection for consistently high or consistently low scores on a behavioural trait might also inadvertently affect the systems that determine the developmental plasticity and/or the contextual plasticity of that trait.

For instance, imagine that we are interested in adults collected from a natural population who score extremely high on a test of ‘aggressiveness’. Based on the principles outlined herein, an extremely high score would only occur under a rare confluence of events: an individual whose genotype supported high levels of aggressiveness, who was raised from conception under conditions that favoured the development of high levels of aggressiveness, and who was confronted with an opponent whose phenotype encouraged a highly aggressive response. However, despite our best efforts, even the most carefully designed laboratory experiments lack complete control over all of the environmental and experiential factors that affect the development and expression of behavioural traits. As a result of uncontrolled variation in any of these factors, an adult who had an extremely high score on one test would be unlikely to have an equally high score on a second test. Similarly, as a result of uncontrolled variation in experiential factors affecting development, a clone or isogenic strain with an extremely high mean score in one generation would be unlikely to produce offspring with an equally high score in the following generation. As a consequence, an individual or a genotype with very high aggressive scores in one test would be likely to have lower scores in a subsequent test. This is an example of the phenomenon called ‘regression to the mean’, discussed in Section II.3*b*.

Now, consider adults from a line that has been subject to strong artificial selection for high aggressiveness for many generations. Such a line is a product of selection for individuals that consistently express extremely high levels of aggression in the face of uncontrolled variation in contextual stimuli (e.g. individuals with low contextual plasticity in aggressive situations), and for genotypes that are consistently highly aggressive in spite of uncontrolled variation in rearing conditions from one generation to the next (i.e. low developmental plasticity for aggressiveness). That is, by choosing individuals or genotypes that are consistently highly

aggressive both within and across generations, investigators may have selected for genes that reduce developmental plasticity and contextual plasticity in aggressive situations. In that situation, strong selection for extremely high 'aggressiveness' would have also selected against mechanisms that down-regulate or modulate the expression of aggressive behaviour in free-living animals.

As was noted earlier, there is already some support for the notion that lines of mice artificially selected for short attack latencies (SAL) are less contextually plastic in aggressive situations than lines selected for long attack latencies (LAL). Indeed, the tendency of SAL males to attack females as well as males is so far from the norm for this species as to be termed 'pathological' (Benus *et al.*, 1990; Caramaschi *et al.*, 2008). Similarly, in male fruit flies, *Drosophila melanogaster*, aggressive behaviour is normally only expressed when males are located on a food resource, and high levels of aggressive behaviour only occur in the presence of cues from adult females (Hoffmann, 1987; Hoffmann & Cacoyianni, 1989). However, following generations of strong artificial selection for high aggressiveness, males from a 'high aggression' line readily engaged in vicious escalated fights in arenas that lacked any food or cues from females (Dierick & Greenspan, 2006).

Other results are consistent with the notion that lines selected for high aggressiveness may be less developmentally plastic than lines selected for low aggressiveness. For instance, variation in experiential factors early in life (handling, rearing conditions) had less effect on the subsequent aggressive behaviour of SAL lines of house mice than that of LAL lines (Benus, 1999; Nyberg *et al.*, 2004), although ceiling and floor effects might have contributed to these results.

Given the widespread use of selected lines as 'model systems' to study the proximate bases of personality traits, we need to know whether these lines differ from non-selected individuals with respect to developmental plasticity or contextual plasticity. In addition, even if the extreme individuals generated by artificial selection were identical to individuals with extreme phenotypes in natural populations, it might be inadvisable to rely too heavily on selected lines for studies of personality development. This is because most individuals in natural populations have intermediate rather than extreme phenotypes. Hence, if there are systematic differences in the development of personality traits for individuals with extreme scores *versus* those with intermediate values (e.g. see Section IV.1), relying exclusively upon individuals at either extremes of the personality spectrum would provide a biased view unrepresentative of the other 90% of the individuals in the population.

(3) Is contextual plasticity a 'meta-personality' trait?

The notion that contextual plasticity might vary across individuals raises the question of whether contextual plasticity might be a personality trait in its own right. For instance, individuals whose responses are more sensitive to subtle differences in contextual stimuli in one functional situation might also be more sensitive to slight differences in contextual stimuli in other situations. This idea has been advanced in

the literature on coping styles in animals (Benus, Koolhaas & Van Oortmerssen, 1987; Koolhaas *et al.*, 1999; reviewed in Sih & Bell, 2008; Wolf *et al.*, 2008), as well as in the human personality literature, which suggests that some individuals may be more sensitive to subtle changes in stimuli than others (Aron & Aron, 1997; Ellis *et al.*, 2006).

Indirect support for this idea comes from comparisons of the SAL and LAL lines of house mice. SAL males consistently attack female as well as male intruders in either their home cage or in novel cages, whereas LAL males attack male but not female intruders (Benus *et al.*, 1990; Caramaschi *et al.*, 2008). Detailed analyses of their aggressive interactions suggested that SAL males are also less sensitive than LAL males to social feedback from conspecifics (Caramaschi *et al.*, 2008), a conclusion bolstered by results indicating that SAL males continue to attack even anaesthetized opponents (Caramaschi *et al.* 2008; Natarajan *et al.*, 2009). Similarly, when males from both lines were tested in a familiar maze, a minor change in the maze had no discernable effect on the behaviour of SAL mice, while the same change elicited immediate responses (pausing, rearing, sniffing, *etc.*) in LAL mice (Benus *et al.*, 1987). Indications of differences in sensitivity to external stimuli extend beyond behavioural traits related to personality, e.g. SAL mice have more difficulty synchronizing their circadian rhythms to changes in daylength than do LAL mice (Benus, Koolhaas & Van Oortmerssen, 1988).

While the results outlined above are consistent with the notion that some individuals might be more contextually plastic in a range of situations than other individuals, it is important to note that the mice discussed in the above paragraphs were the products of artificial selection. As was discussed in Section IV.2, artificial selection itself might lead to changes in the contextual plasticity of personality and other behavioural traits. However, if future studies using unselected individuals indicate that individual differences in contextual plasticity generalize across contexts, then these findings could lead to major changes in the way we think about the organisation of behaviour. Instead of assuming that behaviour is organised into separate modules that correspond to different motivational or functional systems, contextual generality in contextual plasticity would imply that some consistent individual differences in behaviour are so pervasive as to affect the ways that individuals interact with the external world in a wide range of motivational or functional situations. In that case, one could ask about the physiological systems that are related to individual differences in contextual plasticity, about the gene-environment interactions that guide their development, and about the potential costs and benefits of different levels of contextual plasticity under natural conditions. For example, animals that are highly sensitive to subtle cues in the external environment might make more informed decisions than those who are less sensitive to those cues, but might also take longer to make decisions, or require more neural machinery to make them. In addition, consistent individual differences in contextual plasticity might encourage consistent individual differences

in developmental plasticity, if animals that are relatively insensitive to contextual cues are also relatively insensitive to environmental stimuli that lead to developmental changes in personality traits in other members of their species.

(4) How does experience affect correlations between personality and physiological traits?

Correlations between personality traits and physiological traits have long been a focus of research in the literature on coping styles, which consider relationships between consistent individual differences in physiological traits such as the pituitary-adrenal response regulating the production of cortisol and corticosterone or the regulation of (nor)adrenaline, and consistent individual differences in the behaviour that animals express in potentially threatening situations (e.g. fight *versus* flight) or in novel areas (e.g. open field tests) (Koolhaas *et al.*, 1999, 2007). More recently, it has been suggested that correlations across individuals between personality traits (e.g. foraging under predation risk or aggressiveness) and differentially consistent physiological traits (e.g. growth or fecundity) may help account for the differential consistency of those personality traits (Stamps, 2007; Biro & Stamps, 2008). It is often assumed either that correlations between physiological or neurobiological traits and personality traits reflect a causal relationship that is fixed throughout life (i.e. 'genetic'), or that such relationships are only affected by experiential factors that occur very early in life, when the physiological and behavioural systems in question are undergoing rapid development (see also Section IV.5). For example, structural consistency in behavioural stress responses has been interpreted as evidence for temporally consistent individual differences in a general control system that affects impulsivity or sensitivity to environmental stimuli via the serotonergic system (Veenema, Koolhaas & De Kloet, 2004). However, it is an open question whether such links between behavioural and physiological traits are highly resistant to change over the lifetimes of individuals. If further research indicates that correlations between behavioural and physiological traits do change as a result of experiential factors, then experimental studies involving manipulations of those same factors can be used to suggest and test novel hypotheses about the causal mechanisms that are responsible for the formation and maintenance of links between behavioural and physiological traits.

Indeed, recent empirical studies indicate that widespread assumptions about the temporal consistency of links between physiological and personality traits may not be correct. Ruiz-Gomez *et al.* (2008) focused on two lines of rainbow trout artificially selected for high (HR) and low (LR) post-stress plasma cortisol levels. As predicted from the coping style literature, when tested in a novel environment, fish from the LR ('proactive') line began feeding sooner than those from the HR ('reactive') line. Previously it was assumed that the behavioural differences between the LR and HR lines were stable throughout life (Overli *et al.*, 2007), so Ruiz-Gomez *et al.* (2008) were surprised when they tested adult

fish from the HR and LR lines who had been transported to another laboratory, a procedure that involved seven days of starvation, as well as a variety of other stressors. After this (presumably) traumatic event, there was no change in the stress-response profiles of the two lines: HR fish still had higher post-stressor cortisol levels than did LR fish. However, there was a striking reversal in the fishes' relative scores on the feeding test: not only did the fish from the LR line have longer feeding latencies than those from the HR line, but one year later, the LR fish still fed significantly later than the HR fish. Hence, the assumption that relationships between stress physiology and behaviour established early in life are stable thereafter may be premature.

These findings also imply that some of the relationships between personality and physiological traits which appear to be quite stable and predictable may owe some of this predictability and stability to the fact that the experimental subjects are maintained from conception to the time of testing under a single, stable set of conditions. Not only are captive animals typically reared and maintained in conditions that rarely occur in the field (e.g. highly nutritious foods *ad libitum*, no predators or competitors, restricted opportunities for movement, simplified habitat structure), but the same conditions are typically maintained throughout the lives of these animals. In such cases, it is perhaps not surprising that relationships between physiological traits and behavioural traits are highly stable during life. By contrast, a developmental reaction norm perspective suggests that relationships between personality traits and physiological traits are likely to vary as a function of experiential factors that are salient for free-living members of a species. By extension, it argues that experimental manipulations of those factors may provide valuable insights into the proximate mechanisms that contribute to the formation and maintenance (or the lack thereof) of correlations between behavioural and physiological traits in free-living animals.

(5) Can we predict the ages/life stages when experiential factors are likely to affect differential consistency?

Thus far, most studies of the effects of experience on differential consistency have focused on experiences that occur early in life (from conception to the end of parental care), based on the premise that plasticity of the brain and neuroendocrine systems is high at this stage of life. However, even in mammals and birds for which early experience clearly is important, the timing of sensitive periods can vary as a function of experience. In addition, the behaviour that is affected by experiences during a sensitive period may only be expressed at certain ages later in life, and the effects of early experience may be consolidated or overruled by later experiences (Bateson, 1979; Immelman *et al.*, 1991; Hogan & Bolhuis, 2005; Leitner & Catchpole, 2007). As a result, we need to be cautious when assuming that individual differences in personality traits established early in life are necessarily stable thereafter. For instance, at the proximate level, recent studies of rodents indicate that

maternal experiences during gestation or maternal behaviour after birth can have long-lasting effects on the behaviour of their offspring, *via* changes in patterns of DNA methylation (Champagne, 2008). However, although it was formerly assumed that DNA methylation patterns established early in life were subsequently stable, recent data suggest that DNA methylation is potentially responsive to environmental stimuli throughout life (Szyf, McGowan & Meaney, 2008).

Another possible reason why researchers tend to assume that experiences early in life have the strongest impact on the development of behavioural and other phenotypic traits (e.g. Crews, 2008; Nijland, Ford & Nathanielsz, 2008) is that in the laboratory, experimental subjects are typically maintained under constant, benign conditions following the experimental perturbation that was the focus of the study. In the absence of highly salient experiences later in life (e.g. attacks from predators or conspecifics, periods of starvation, changes in social group size or composition), there would be little reason to expect individual differences established early in life to change later in life (see also Section IV.4).

Recently, we have begun to see studies which suggest that salient experiences that occur after animals have become independent of their parents can affect the differential consistency of personality traits. For instance, juvenile rainbow trout can be categorized as 'shy' and 'bold' based on their scores on a test involving latency to approach a novel object, where the 'shy' fish have longer approach latencies than the 'bold' fish. Frost *et al.* (2007) asked whether particular types of experience (e.g. observation of a 'shy' conspecific demonstrator in an adjacent tank over a four-day period) affected the subject's relative scores on the same test conducted after exposure to the demonstrator. After observing a 'shy' demonstrator, the previously 'shy' fish had a lower average latency to approach the novel object than did the previously 'bold' fish, while bold fish after the same experience increased their latency, implying that in this species, social experiences at the juvenile age/stage may affect the differential consistency of personality traits. In another experiment, exposing subadult three-spined stickleback to predatory trout changed the relationship between their scores in tests measuring 'boldness' and 'aggressiveness' (Bell & Sih, 2007). Among the subjects that survived this experience, prior to the exposure to predators, the relationship between boldness and aggressiveness was weak ($r = 0.18$), but after the exposure the effect size for this same relationship had increased to $r = 0.46$. Further analyses revealed that exposure to predators had reduced the differential consistency of aggressiveness, i.e. there was no discernable relationship between aggressiveness scores of the same individuals before and after exposure to predators. However, the same experience had no discernable effect on the differential consistency of boldness. As a result, individuals who were relatively bold (but not necessarily also aggressive) prior to the experience tended to be both bold and aggressive after it. These examples suggest that in fish, at least, salient experiences that occur months after individuals have begun

living independent lives can have pronounced effects on both differential consistency and structural consistency.

(6) How does spatial and temporal variation in environmental factors affect personality development?

In contrast to the situation in the laboratory, where subjects typically develop under one set of environmental conditions for an extended period of time, in nature individuals are likely to develop in areas in which environmental conditions vary across time and space. From a developmental perspective, 'temporal' refers to the environmental conditions experienced by an individual between conception and death, while 'spatial' refers to environmental conditions in the areas that are within the dispersal distances that an individual might travel over the course of its lifetime. Here, we suggest that temporal and spatial variation at the scales that are relevant to individual development are likely to have different effects on individual stability and differential consistency of personality traits.

First imagine a situation in which environmental factors affecting personality development vary temporally but not spatially. If individuals are developmentally plastic, then temporal variation in these environmental factors would be expected to reduce individual stability in scores for personality traits. Further, if individuals within the same population differ with respect to their developmental plasticity (variation among individuals in developmental reaction norms), then temporal variation in environmental factors would also reduce differential consistency. Hence, temporal variation in environmental factors that affect personality development would be likely to reduce temporal consistency of personality traits at both the individual level (individual stability) and at the group level (differential consistency).

Second, imagine a different situation, in which environmental factors that affect personality development are constant across the lifetimes of each cohort, but vary across distances that can be readily traveled by individuals of that species. If individuals differ with respect to the environments in which they prefer to live and develop, then as a result of niche-picking or niche-construction, different individuals in the same population might be able to maintain themselves in consistently different environments for extended periods of time. In this situation, individual stability and differential consistency might be higher than if all of the individuals in the population had been forced to live in one of these environments, or if all of the individuals in the population randomly drifted from one environment to another over the course of their lives.

This line of reasoning suggests that differences between the laboratory and the field in the stability and consistency of personality traits may depend on whether environmental variation in nature is primarily temporal or primarily spatial. If the variation is primarily temporal, then one would expect personality to be more stable for individuals raised under constant conditions in the laboratory than for otherwise comparable individuals who developed in the

field. Conversely, if the variation is primarily spatial, we might expect the reverse to be true.

(7) When during ontogeny are personality traits likely to be temporally unstable?

Intuition suggests that the differential and structural consistency of personality traits might vary over the course of a lifetime, and descriptive studies confirm that this is true for humans (Roberts & DelVecchio, 2000; Roberts *et al.*, 2001; Caspi *et al.*, 2005). There are at least two reasons why personality traits that are temporally stable at a given age or life stage might be unstable earlier or later in life. First, if personality traits are linked at the proximate level to underlying physiological processes such as growth or metabolic rate (Stamps, 2007; Biro & Stamps, 2008; Careau *et al.*, 2008), we might expect instability in personality traits to occur when the physiological systems to which they are linked are undergoing major changes. For instance, personality traits might become unstable during periods of rapid morphogenesis, metamorphosis, sexual maturation or other stages of ontogeny when physiological and morphological systems are undergoing major reorganisation. In contrast, ‘external’ hypotheses for personality instability rest on the assumption that individuals are developmentally plastic, and suggest that changes in personality within individuals are most likely to occur when those individuals shift from one physical or social environment to another, e.g., as a result of natal or breeding dispersal, migration, or recruitment to a new social group or neighborhood. Obviously, as is the case for genetic *versus* experiential influences on behavioural development, these two hypotheses are neither mutually exclusive nor independent.

Testing these hypotheses requires types of data that are rarely collected in the same study, namely, information about the timing of changes in physiological systems linked to personality traits, and information about the timing of changes in social and physical environments for free-living members of the species. For instance, longitudinal studies of three-spined stickleback and dumpling squid conducted in the laboratory suggest that differential consistency in personality traits declines around the time of sexual maturity (Bell & Stamps, 2004; Sinn *et al.*, 2008), but for neither of these species do we know whether free-living individuals change their physical or social habitats when they mature. In female brush mice, *Peromyscus boylii*, natal dispersal and sexual maturity occur at the same age (Mabry & Stamps, 2008), so for this species, both hypotheses predict lower individual stability and differential consistency in personality traits for dispersal-age animals than at younger ages (when juveniles are living with their mother and siblings in the natal home range) or at older ages (when mature animals are living in a new neighborhood). Discriminating between the two hypotheses would be easier in species in which major reorganisation of physiological systems occurs at different ages or life stages than when animals transfer to novel physical and/or social environments. For instance, in humans in western cultures, sexual maturation (puberty)

typically occurs many years before individuals leave home and establish themselves in a new area and form new social relationships. Studies of personality development in these populations indicate personality traits are relatively stable through puberty, but become unstable during young adulthood, a time of major transitions in physical and social environments (Caspi *et al.* 2005).

(8) How can we simultaneously analyse contextual and temporal components of personality?

As was noted above, although in principle contextual reaction norms are based on scores for behaviour expressed in different contexts at the same age and period of time, in practice, descriptions of the contextual reaction norms of individuals are necessarily based on tests conducted in different contexts at different times. As a result, variation among individuals in the temporal stability of behaviour across the testing period will affect estimates of their contextual reaction norms and contextual plasticity. Those studying contextual reaction norms can not simply assume that individual stability is high across the testing period for every behaviour of every individual in every context, or that temporal changes in behaviour in one or more contexts are comparable for all of the individuals in the sample. In addition, biologists who use particular statistical models to study contextual reaction norms need to be aware of the assumptions those models are making with respect to individual differences in temporal changes in the traits of interest. For instance, Nussey *et al.* (2007) showed how random regression models can be used to study contextual plasticity, but their procedure assumes either that an individual’s scores for the behaviour expressed in each context do not change over the sampling period, or that the behaviour expressed in each context changes the same way for every individual in a sample as a function of age or time.

Thus far, no one working on animal personality has attempted to simultaneously analyze the contextual plasticity and individual stability of the same set of subjects. Developing experimental protocols and statistical tools to address this problem is a high priority, because as we have emphasized throughout this article, personality involves both contextual and temporal consistency, both of which can vary among the individuals in the same population. In addition, more comprehensive methods for the analysis of personality would allow us to answer new and important questions. For instance, we could ask whether individual differences in variation in behaviour across contexts is correlated with individual differences in variation in behaviour across time, i.e. whether contextual plasticity is correlated to individual stability across individuals. This latter question is relevant to current discussions about variation among individuals in their degree of ‘specialization’ (e.g. Sih & Bell, 2008; Reale & Dingemans, 2009), in asking whether animals who are highly specialized with respect to their behaviour at a given time (low contextual plasticity) are also more likely to maintain the same behaviour in a given context over time (high individual stability).

V. CONCLUSIONS

- (1) Although relatively few investigators have explicitly studied the development of animal personality, understanding the functional significance, evolution and causation of personality is likely to be enriched by understanding its development. Changes in personality across ontogeny may shed light on age or stage-specific adaptations to environmental conditions that vary across the lifespan, and provide valuable insights into the causal mechanisms and selective pressures that are responsible for formation and maintenance of personality. However, an ontogenetic approach requires familiarity with a variety of concepts that are ignored in 'snapshot' studies that focus on personality at a given age or time; outlining and explaining these concepts has been a major goal of this review.
- (2) Understanding the development of personality requires studies of behaviour over longer periods of time, and under different sets of environmental conditions, than is typically the case for animal personality studies. Currently, estimates of the differential and structural consistency of personality traits are usually based on relatively short inter-test intervals, and there is sometimes a tendency to assume that traits that are temporally stable for short periods of time are also stable over much longer periods, including across the entire lifetime. In addition, many studies of animal personality are conducted under constant, benign, and highly simplified conditions in the laboratory, even though their goal is to understand the causes and consequences of personality for animals that live in complex, spatially and temporally variable environments in nature.

We need more studies that carefully map the development of personality traits from early life onwards, either in the field or under semi-natural conditions that are based, as far as possible, on the natural life history of the species in question. Long-term, longitudinal studies of personality development can reveal whether traits or correlations among traits are consistent or inconsistent at different ages or life stages. In addition, personality development for individuals living in the field or in semi-natural conditions can be compared to personality development for members of the same cohort or selected lines reared in standardized laboratory conditions. This sort of comparison would help establish the validity of some of the methods currently used to study animal personality (e.g. the use of artificially selected lines as model systems), and shed light on the extent to which niche-picking or niche-construction might contribute to the formation and maintenance of personality in free-living animals.

- (3) Contextual reaction norms and developmental reaction norms provide a framework for addressing some of the most important and interesting questions about

animal personality, e.g. the extent to which contextual plasticity varies among individuals and genotypes, how interactions and correlations between genes and experiential factors contribute to the production and maintenance of personality, and how behavioural processes such as niche-picking and niche-construction affect the establishment and maintenance of personality in natural populations. In combination, contextual and developmental reaction norms can provide powerful insights into the ways that genes, experiential factors, and correlations and interactions between genes and experiential factors contribute to the individual differences in behaviour, contextual consistency and temporal consistency that are the hallmarks of animal personality.

- (4) Although this review focuses on behavioural traits that have attracted the most attention in the animal personality and related literatures (e.g. aggressiveness, boldness, activity, *etc.*, see Reale *et al.*, 2007; Sih & Bell, 2008), the concepts outlined herein apply equally well to the many other behavioural and physiological traits that are differentially consistent and whose expression varies as a function of contextual stimuli (e.g. consistent individual differences in parenting styles or physiological responses to 'stressful' stimuli). Hence, studies of personality development encourage us to consider general questions about the organisation and development of the many phenotypic traits that vary over the short term as a function of contextual cues at the time of trait expression, and over the long term as a function of experiential and environmental factors that occurred earlier in the lifetime of that individual.
- (5) We hope to encourage future studies of the development of personality, by outlining the concepts that need to be considered in empirical and theoretical studies of this topic, and by offering a list of questions about personality development that would benefit from attention from empiricists. In addition, we hope that this review motivates theoreticians and empiricists who are interested in the causation, functional significance or evolution of personality to identify and reconsider the assumptions about development that they are making, either explicitly or unwittingly, in their own studies.

VI. ACKNOWLEDGEMENTS

We are very grateful to Alison Bell, Peter Biro, Niels Dinglemanse and two anonymous reviewers for their comments, advice and suggestions on previous versions of this manuscript.

VII. REFERENCES

- ANHOLT, R. R. H. (2004). Genetic modules and networks for behavior: lessons from *Drosophila*. *Bioessays* **26**, 1299–1306.

- ARNOLD, K. E., RAMSAY, S. L., DONALDSON, C. & ADAM, A. (2007). Parental prey selection affects risk-taking behaviour and spatial learning in avian offspring. *Proceedings of the Royal Society B-Biological Sciences* **274**, 2563–2569.
- ARON, E. N. & ARON, A. (1997). Sensory-processing sensitivity and its relation to introversion and emotionality. *Journal of Personality and Social Psychology* **73**, 345–368.
- ASENDORPF, J. B. (1990). The measurement of individual consistency. *Methodika* **4**, 1–23.
- ASENDORPF, J. B. (1992). Beyond stability – predicting interindividual differences in intraindividual change. *European Journal of Personality* **6**, 103–117.
- BAERENDS, G. P. (1976). Functional organization of behavior. *Animal Behaviour* **24**, 726–738.
- BATESON, P. (1979). How do sensitive periods arise and what are they for. *Animal Behaviour* **27**, 470–486.
- BELL, A. M. (2007). Future directions in behavioural syndromes research. *Proceedings of the Royal Society B-Biological Sciences* **274**, 755–761.
- BELL, A. M., HANKISON, S. J. & LASKOWSKI, K. L. (2009). The repeatability of behaviour: a meta-analysis. *Animal Behaviour*, **77** 771–783.
- BELL, A. M. & SIH, A. (2007). Exposure to predation generates personality in threespined sticklebacks (*Gasterosteus aculeatus*). *Ecology Letters* **10**, 828–834.
- BELL, A. M. & STAMPS, J. A. (2004). Development of behavioural differences between individuals and populations of sticklebacks, *Gasterosteus aculeatus*. *Animal Behaviour* **68**, 1339–1348.
- BENUS, R. F. (1999). Differential effect of handling on adult aggression in male mice bidirectionally selected for attack latency. *Aggressive Behavior* **25**, 365–368.
- BENUS, R. F., DEN DAAS, S. J., KOOLHAAS, J. M. & VAN OORTMERSSEN, G. A. (1990). Routine formation and flexibility in social and nonsocial behavior of aggressive and non-aggressive male mice. *Behaviour* **112**.
- BENUS, R. F., KOOLHAAS, J. M. & VAN OORTMERSSEN, G. A. (1987). Individual-differences in behavioral reaction to a changing environment in mice and rats, pp. 105–122.
- BENUS, R. F., KOOLHAAS, J. M. & VAN OORTMERSSEN, G. A. (1988). Aggression and adaptation to the light-dark cycle – role of intrinsic and extrinsic control. *Physiology & Behavior* **43**, 131–137.
- BIRO, P. A. & STAMPS, J. A. (2008). Are animal personality traits linked to life-history productivity? *Trends in Ecology and Evolution*.
- BLAND, J. M. & ALTMAN, D. G. (1994). Some examples of regression towards the mean VII. *British Medical Journal* **309**, 780–780.
- BOON, A. K., REALE, D. & BOUTIN, S. (2007). The interaction between personality, offspring fitness and food abundance in North American red squirrels. *Ecology Letters* **10**, 1094–1104.
- BOON, A. K., REALE, D. & BOUTIN, S. (2008). Personality, habitat use, and their consequences for survival in North American red squirrels *Tamiasciurus hudsonicus*. *Oikos* **117**, 1321–1328.
- BOSSDORF, O., RICHARDS, C. L. & PIGLIUCCI, M. (2008). Epigenetics for ecologists. *Ecology Letters* **11**, 106–115.
- BRIFFA, M., RUNDLE, S. D. & FRYER, A. (2008). Comparing the strength of behavioural plasticity and consistency across situations: animal personalities in the hermit crab *Pagurus bernhardus*. *Proceedings of the Royal Society B-Biological Sciences* **275**, 1305–1311.
- BRODIE, E. D. & RUSSELL, N. H. (1999). The consistency of individual differences in behaviour: temperature effects on antipredator behaviour in garter snakes. *Animal Behaviour* **57**, 445–451.
- BURLEY, N. (1981). Sex-ratio manipulation and selection for attractiveness. *Science* **211**, 721–722.
- BUSS, A. H. & PLOMIN, R. (1984). *Temperament: Early developing personality traits*. Lawrence Erlbaum Associates, Hillsdale, New York.
- CARAMASCHI, D., DE BOER, S. F., DE VRIES, H. & KOOLHAAS, J. M. (2008). Development of violence in mice through repeated victory along with changes in Caprefrontal cortex neurochemistry. *Behavioural Brain Research* **189**, 263–272.
- CAREAU, V., THOMAS, D., HUMPHRIES, M. M. & REALE, D. (2008). Energy metabolism and animal personality. *Oikos* **117**, 641–653.
- CARERE, C., DRENT, P. J., KOOLHAAS, J. M. & GROOTHUIS, T. G. G. (2005a). Epigenetic effects on personality traits: early food provisioning and sibling competition, pp. 1329–1355.
- CARERE, C., DRENT, P. J., PRIVITERA, L., KOOLHAAS, J. M. & GROOTHUIS, T. G. G. (2005b). Personalities in great tits, *Parus major*: stability and consistency. *Animal Behaviour* **70**, 795–805.
- CASPI, A. & ROBERTS, B. W. (1999). Personality continuity and change across the life course. In *Handbook of personality: theory and research*. Second Edition (ed. L.A. Pervin & O.P. John), pp 300–326. Guilford Press, New York.
- CASPI, A. & ROBERTS, B. W. (2001). Personality development across the life course: The argument for change and continuity. *Psychological Inquiry* **12**, 49–66.
- CASPI, A., ROBERTS, B. W. & SHINER, R. L. (2005). Personality development: Stability and change. *Annual Review of Psychology* **56**, 453–484.
- CHAMPAGNE, F. A. (2008). Epigenetic mechanisms and the transgenerational effects of maternal care. *Frontiers in Neuroendocrinology* **29**, 386–397.
- CLARK, M. M. & GALEF, B. G. (1995). Prenatal influences on reproductive life-history strategies. *Trends in Ecology & Evolution* **10**, 151–153.
- COLEMAN, K. & WILSON, D. S. (1998). Shyness and boldness in pumpkinseed sunfish: individual differences are context-specific. *Animal Behaviour* **56**, 927–936.
- CORDIER, S. (2008). Evidence for a role of paternal exposures in developmental toxicity. *Basic & Clinical Pharmacology & Toxicology* **102**, 176–181.
- COTE, J. & CLOBERT, J. (2007). Social personalities influence natal dispersal in a lizard. *Proceedings of the Royal Society B-Biological Sciences* **274**, 383–390.
- CREWS, D. (2008). Epigenetics and its implications for behavioral neuroendocrinology. *Frontiers in Neuroendocrinology* **29**, 344–357.
- CREWS, D., GORE, A. C., HSU, T. S., DANGLEBEN, N. L., SPINETTA, M., SCHALLERT, T., ANWAY, M. D. & SKINNER, M. K. (2007). Transgenerational epigenetic imprints on mate preference. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 5942–5946.
- CURLEY, J. P., CHAMPAGNE, F. A., BATESON, P. & KEVERNE, E. B. (2008). Transgenerational effects of impaired maternal care on behaviour of offspring and grandoffspring. *Animal Behaviour* **75**, 1551–1561.

- DALL, S. R. X., HOUSTON, A. I. & MCNAMARA, J. M. (2004). The behavioural ecology of personality: consistent individual differences from an adaptive perspective. *Ecology Letters* **7**, 734–739.
- DE FRUYT, F., BARTELS, M., VAN LEEUWEN, K. G., DE CLERCQ, B., DECUYPER, M. & MERVIELDE, I. (2006). Five types of personality continuity in childhood and adolescence. *Journal of Personality and Social Psychology* **91**, 538–552.
- DIERICK, H. A. & GREENSPAN, R. J. (2006). Molecular analysis of flies selected for aggressive behavior. *Nature Genetics* **38**, 1023–1031.
- DINGEMANSE, N. J. (2007). An evolutionary ecologist's view of how to study the persistence of genetic variation in personality. *European Journal of Personality* **21**, 593–596.
- 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 of London Series B-Biological Sciences* **271**, 847–852.
- DINGEMANSE, N. J., BOTH, C., DRENT, P. J., VAN OERS, K. & VAN NOORDWIJK, A. J. (2002). Repeatability and heritability of exploratory behaviour in great tits from the wild. *Animal Behaviour* **64**, 929–938.
- 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**, 1128–1138.
- DRENT, P. J., VAN OERS, K. & VAN NOORDWIJK, A. J. (2003). Realized heritability of personalities in the great tit (*Parus major*). *Proceedings of the Royal Society of London Series B-Biological Sciences* **270**, 45–51.
- DRUMMOND, H. (2006). Dominance in vertebrate broods and litters. *Quarterly Review of Biology* **81**, 3–32.
- ELLIS, B. J., JACKSON, J. J. & BOYCE, W. T. (2006). The stress response systems: Universality and adaptive individual differences. *Developmental Review* **26**, 175–212.
- FALCONER, D. S. & MACKAY, T. F. (1996). *Introduction to Quantitative Genetics*, 4th Edition edition. Longman, New York.
- FEAVER, J., MENDEL, M. & BATESON, P. (1986). A method for rating the individual distinctiveness of domestic cats. *Animal Behaviour* **34**, 1016–1025.
- FIDLER, A. E., VAN OERS, K., DRENT, P. J., KUHN, S., MUELLER, J. C. & KEMPENAERS, B. (2007). Drd4 gene polymorphisms are associated with personality variation in a passerine bird. *Proceedings of the Royal Society B-Biological Sciences* **274**, 1685–1691.
- FLEESON, W. (2004). Moving personality beyond the person-situation debate – The challenge and the opportunity of within-person variability. *Current Directions in Psychological Science* **13**, 83–87.
- 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**, 235–262.
- FROST, A. J., WINROW-GIFFEN, A., ASHLEY, P. J. & SNEDDON, L. U. (2007). Plasticity in animal personality traits: does prior experience alter the degree of boldness? *Proceedings of the Royal Society B-Biological Sciences* **274**, 333–339.
- FULLER, T., SARKAR, S. & CREWS, D. (2005). The use of norms of reaction to analyze genotypic and environmental influences on behavior in mice and rats. *Neuroscience and Biobehavioral Reviews* **29**, 445–456.
- GERLAI, R. & CSANYI, V. (1990). Genotype-environment interaction and the correlation structure of behavioral elements in Paradise Fish (*Macropodus opercularis*). *Physiology & Behavior* **47**, 343–356.
- GIBERT, P., MORETEAU, B., DAVID, J. R. & SCHEINER, S. M. (1998). Describing the evolution of reaction norm shape: Body pigmentation in *Drosophila*. *Evolution* **52**, 1501–1506.
- GIL, D. & FAURE, J. M. (2007). Correlated response in yolk testosterone levels following divergent genetic selection for social behaviour in Japanese quail. *Journal of Experimental Zoology Part a-Ecological Genetics and Physiology* **307A**, 91–94.
- GIL, D., GRAVES, J., HAZON, N. & WELLS, A. (1999). Male attractiveness and differential testosterone investment in zebra finch eggs. *Science* **286**, 126–128.
- GOSLING, S. D. (2001). From mice to men: What can we learn about personality from animal research? *Psychological Bulletin* **127**, 45–86.
- GOTTHARD, K. & NYLIN, S. (1995). Adaptive plasticity and plasticity as an adaptation – a selective review of plasticity in animal morphology and life-history. *Oikos* **74**, 3–17.
- GROOTHUIS, T. G. G. & CARERE, C. (2005). Avian personalities: characterization and epigenesis. *Neuroscience and Biobehavioral Reviews* **29**, 137–150.
- GROOTHUIS, T. G. G., CARERE, C., LIPAR, J., DRENT, P. J. & SCHWABL, H. (2008). Selection on personality in a songbird affects maternal hormone levels tuned to its effect on timing of reproduction. *Biology Letters* **4**, 465–467.
- GROOTHUIS, T. G. G., MULLER, W., VON ENGELHARDT, N., CARERE, C. & EISING, C. (2005). Maternal hormones as a tool to adjust offspring phenotype in avian species. *Neuroscience and Biobehavioral Reviews* **29**, 329–352.
- HAYES, J. P. & JENKINS, S. H. (1997). Individual variation in mammals. *Journal of Mammalogy* **78**, 274–293.
- HOFFMANN, A. A. (1987). A laboratory study of male territoriality in the sibling species *Drosophila melanogaster* and *Drosophila simulans*. *Animal Behaviour* **35**, 807–818.
- HOFFMANN, A. A. & CACOYIANNI, Z. (1989). Selection for territoriality in *Drosophila melanogaster* – correlated responses in mating success and other fitness components. *Animal Behaviour* **38**, 23–34.
- HOFMANN, H. A. (2003). Functional genomics of neural and behavioral plasticity. *Journal of Neurobiology* **54**, 272–282.
- HOGAN, J. A. & BOLHUIS, J. J. (2005). The development of behaviour: trends since Tinbergen (1963). *Animal Biology* **55**, 371–398.
- HUNTINGFORD, F. A. (1976). Relationship between anti-predator behavior and aggression among conspecifics in 3-spined stickleback, *Gasterosteus aculeatus*. *Animal Behaviour* **24**, 245–260.
- IMMELMANN, K., PROVE, R., LASSEK, R. & BISCHOF, H. J. (1991). Influence of adult courtship experience on the development of sexual preferences in Zebra Finch males. *Animal Behaviour* **42**, 83–89.
- IZEM, R. & KINGSOLVER, J. G. (2005). Variation in continuous reaction norms: Quantifying directions of biological interest. *American Naturalist* **166**, 277–289.
- JABLONKA, E. (2007). The development construction of heredity. *Developmental Psychobiology* **49**, 808–817.
- JACOBSON, N. S. & TRAU, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology* **59**, 12–19.

- JANG, K. L. (2007). Behaviour genetics' neglected twin: Gene-environment correlation. *European Journal of Personality* **21**, 604–605.
- JOHNSON, J. C. & SIH, A. (2005). Precopulatory sexual cannibalism in fishing spiders (*Dolomedes triton*): a role for behavioral syndromes. *Behavioral Ecology and Sociobiology* **58**, 390–396.
- JOHNSON, J. C. & SIH, A. (2007). Fear, food, sex and parental care: a syndrome of boldness in the fishing spider, *Dolomedes triton*. *Animal Behaviour* **74**, 1131–1138.
- KAGAN, J., REZNICK, J. S. & SNIDMAN, N. (1988). Biological bases of childhood shyness. *Science* **240**, 167–171.
- KENDLER, K. S. & GREENSPAN, R. J. (2006). The nature of genetic influences on behavior: Lessons from “simpler” organisms. *American Journal of Psychiatry* **163**, 1683–1694.
- KERR, M., LAMBERT, W. W., STATTIN, H. & KLACKENBERGLARSSON, I. (1994). Stability of inhibition in a Swedish longitudinal sample. *Child Development* **65**, 138–146.
- KINGSOLVER, J. G., SHLICHTA, J. G., RAGLAND, G. J. & MASSIE, K. R. (2006). Thermal reaction norms for caterpillar growth depend on diet. *Evolutionary Ecology Research* **8**, 703–715.
- KIRKPATRICK, M. & HECKMAN, N. (1989). A quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters. *Journal of Mathematical Biology* **27**, 429–450.
- KOOLHAAS, J. M., DE BOER, S. F., BUWALDA, B. & VAN REENEN, K. (2007). Individual variation in coping with stress: A multidimensional approach of ultimate and proximate mechanisms. *Brain Behavior and Evolution* **70**, 218–226.
- KOOLHAAS, J. M., KORTE, S. M., DE BOER, S. F., VAN DER VEGT, B. J., VAN REENEN, C. G., HOPSTER, H., DE JONG, I. C., RUIS, M. A. W. & BLOKHUIS, H. J. (1999). Coping styles in animals: current status in behavior and stress-physiology. *Neuroscience and Biobehavioral Reviews* **23**, 925–935.
- KUTSUKAKE, N. & CLUTTON-BROCK, T. H. (2008). Do meerkats engage in conflict management following aggression? Reconciliation, submission and avoidance. *Animal Behaviour* **75**, 1441–1453.
- LEE, J. S. F. & BEREJIKIAN, B. A. (2008). Stability of behavioral syndromes but plasticity in individual behavior: consequences for rockfish stock enhancement. *Environmental Biology of Fishes* **82**, 179–186.
- LEITNER, S. & CATCHPOLE, C. K. (2007). Song and brain development in canaries raised under different conditions of acoustic and social isolation over two years. *Developmental Neurobiology* **67**, 1478–1487.
- LESSELLS, C. M. & BOAG, P. T. (1987). Unrepeatable repeatabilities – a common mistake. *Auk* **104**, 116–121.
- LEWONTIN, R. C. (1974). The analysis of variance and the analysis of causes. *American Journal of Human Genetics* **26**, 400–411.
- LOVE, M. S. & JOHNSON, K. (1998). Aspects of the life histories of grass rockfish, *Sebastes rastrelliger*, and brown rockfish, *S. auriculatus*, from southern California. *California Fisheries Bulletin* **87**, 100–109.
- MABRY, K. E. & STAMPS, J. A. (2008). Dispersing brush mice prefer habitat like home. *Proceedings of the Royal Society B-Biological Sciences* **275**, 543–548.
- MAGNHAGEN, C. (2007). Social influence on the correlation between behaviours in young-of-the-year perch. *Behavioral Ecology and Sociobiology* **61**, 525–531.
- MALMKVIST, J. & HANSEN, S. W. (2002). Generalization of fear in farm mink, *Mustela vison*, genetically selected for behaviour towards humans. *Animal Behaviour* **64**, 487–501.
- MARTIN, J. G. A. & REALE, D. (2008). Temperament, risk assessment and habituation to novelty in eastern chipmunks, *Tamias striatus*. *Animal Behaviour* **75**, 309–318.
- MCNAMARA, J. M., STEPHENS, P. A., DALL, S. R. X. & HOUSTON, A. I. (2009). Evolution of trust and trustworthiness: social awareness favours personality differences. *Proceedings of the Royal Society B-Biological Sciences* **276**, 605–613.
- MCPHEE, M. V. & QUINN, T. P. (1998). Factors affecting the duration of nest defense and reproductive lifespan of female sockeye salmon, *Oncorhynchus nerka*. *Environmental Biology of Fishes* **51**, 369–375.
- MENDL, M. & PAUL, E. S. (1991). Parental care, sibling relationships and the development of aggressive behavior in 2 lines of wild House Mice, pp. 11–41.
- MOORE, A. J., WOLF, J. B. & BRODIE III, E. D. (1998). The influence of direct and indirect genetic effects on the evolution of behaviour: social and sexual selection meet maternal effects. In *Maternal effects as adaptations* (ed. T. A. Mousseau and C. A. Fox), pp. 22–41. Oxford University Press, Oxford.
- MUNAFÒ, M. R., CLARK, T. G., MOORE, L. R., PAYNE, E., WALTON, R. & FLINT, J. (2003). Genetic polymorphisms and personality in healthy adults: A systematic review and meta-analysis. *Molecular Psychiatry* **8**, 471–484.
- NATARAJAN, D., DE VRIES, H., SAALTINK, D. J., DE BOER, S. F. & KOOLHAAS, J. (2009). Delineation of violence from functional aggression in Mice: An ethological approach. *Behavior Genetics* **39**, 73–90.
- NATOLI, E., SAY, L., CAFAZZO, S., BONANNI, R., SCHMID, M. & PONTIER, D. (2005). Bold attitude makes male urban feral domestic cats more vulnerable to Feline Immunodeficiency Virus. *Neuroscience and Biobehavioral Reviews* **29**, 151–157.
- NELSON, R. J. (2005). *An Introduction to Behavioral Neurobiology*, Third edition. Sinauer, Sunderland, MA.
- NETTLE, D. (2006). The evolution of personality variation in humans and other animals. *American Psychologist* **61**, 622–631.
- NIJLAND, M. J., FORD, S. P. & NATHANIELSZ, P. W. (2008). Prenatal origins of adult disease. *Current Opinion in Obstetrics & Gynecology* **20**, 132–138.
- NUSSEY, D. H., WILSON, A. J. & BROMMER, J. E. (2007). The evolutionary ecology of individual phenotypic plasticity in wild populations. *Journal of Evolutionary Biology* **20**, 831–844.
- NYBERG, J., SANDNABBA, K., SCHALKWYK, L. & SLUYTER, F. (2004). Genetic and environmental (inter)actions in male mouse lines selected for aggressive and nonaggressive behavior. *Genes Brain and Behavior* **3**, 101–109.
- OVERLI, O., SORENSEN, C., PULMAN, K. G. T., POTTINGER, T. G., KORZAN, W., SUMMERS, C. H. & NILSSON, G. E. (2007). Evolutionary background for stress-coping styles: Relationships between physiological, behavioral, and cognitive traits in non-mammalian vertebrates. *Neuroscience and Biobehavioral Reviews* **31**, 396–412.
- OVERLI, O., WINBERG, S. & POTTINGER, T. G. (2005). Behavioral and neuroendocrine correlates of selection for stress responsiveness in rainbow trout – a review. *Integrative and Comparative Biology* **45**, 463–474.
- PIGLIUCCI, M. (2001). *Phenotypic plasticity*. John Hopkins University Press, Baltimore.

- PITKOW, L. J., SHARER, C. A., REN, X. L., INSEL, T. R., TERWILLIGER, E. F. & YOUNG, L. J. (2001). Facilitation of affiliation and pair-bond formation by vasopressin receptor gene transfer into the ventral forebrain of a monogamous vole. *Journal of Neuroscience* **21**, 7392–7396.
- PLOMIN, R., DEFRIES, J. C. & LOEHLIN, J. C. (1977). Genotype-environment interaction and correlation in analysis of human behavior. *Psychological Bulletin* **84**, 309–322.
- REALE, D. & DINGEMANSE, N. J. in press. Personality and individual social specialization. In *Social behavior: genes, ecology and evolution* (ed. T. Szekely, A. Moore and J. Komdeur).
- REALE, D. & FESTA-BIANCHET, M. (2003). Predator-induced natural selection on temperament in bighorn ewes. *Animal Behaviour* **65**, 463–470.
- REALE, D., GALLANT, B. Y., LEBLANC, M. & FESTA-BIANCHET, M. (2000). Consistency of temperament in bighorn ewes and correlates with behaviour and life history. *Animal Behaviour* **60**, 589–597.
- REALE, D., READER, S. M., SOL, D., MCDUGALL, P. T. & DINGEMANSE, N. J. (2007). Integrating animal temperament within ecology and evolution. *Biological Reviews* **82**, 291–318.
- RENAULT, J., GHEUSI, G. & AUBERT, A. (2008). Changes in social exploration of a lipopolysaccharides-treated conspecific in mice: Role of environmental cues. *Brain Behavior and Immunity* **22**, 1201–1207.
- RICHARD, F. J., AUBERT, A. & GROZINGER, C. M. (2008). Modulation of social interactions by immune stimulation in honey bee, *Apis mellifera*, workers. *Bmc Biology* **6**, .
- ROBERTS, B. W., CASPI, A. & MOFFITT, T. E. (2001). The kids are alright: Growth and stability in personality development from adolescence to adulthood. *Journal of Personality and Social Psychology* **81**, 670–683.
- ROBERTS, B. W. & DELVECCHIO, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin* **126**, 3–25.
- ROBSON, S. L. & WOOD, B. (2008). Hominid life history: reconstruction and evolution. *Journal of Anatomy* **212**, 394–425.
- ROGERS, J., SHELTON, S. E., SHELEDY, W., GARCIA, R. & KALIN, N. H. (2008). Genetic influences on behavioral inhibition and anxiety in juvenile rhesus macaques. *Genes Brain and Behavior* **7**, 463–469.
- RUIZ-GOMEZ, M. D., KITTELSEN, S., HOGLUND, E., HUNTINGFORD, F. A., SORENSEN, C., POTTINGER, T. G., BAKKEN, M., WINBERG, S., KORZAN, W. J. & OVERLI, O. (2008). Behavioral plasticity in rainbow trout (*Oncorhynchus mykiss*) with divergent coping styles: When doves become hawks. *Hormones and Behavior* **54**, 534–538.
- RUTTER, M. & SILBERG, J. (2002). Gene-environment interplay in relation to emotional and behavioral disturbance. *Annual Review of Psychology* **53**, 463–490.
- SAETRE, P., STRANDBERG, E., SUNDGREN, P. E., PETTERSSON, U., JAZIN, E. & BERGSTROM, T. F. (2006). The genetic contribution to canine personality. *Genes Brain and Behavior* **5**, 240–248.
- SANSON, A., PEDLOW, R., CANN, W., PRIOR, M. & OBERKLAID, F. (1996). Shyness ratings: Stability and correlates in early childhood. *International Journal of Behavioral Development* **19**, 705–724.
- SCARR, S. & MCCARTNEY, K. (1983). How people make their own environments – A theory of genotype-environment effects. *Child Development* **54**, 424–435.
- SCHAEFFER, L. R. (2004). Application of random regression models in animal breeding. *Livestock Production Science* **86**, 35–45.
- SCHLICHTING, C. D. & PIGLIUCCI, M. (1998). *Phenotypic evolution: a reaction norm perspective*. Sinauer, Sunderland, MA.
- SCHMALHAUSEN, I. I. (1949). *Factors of Evolution*. Blakiston, Philadelphia, PA.
- SCHUERGER, J. M., ZARRELLA, K. L. & HOTZ, A. S. (1989). Factors that influence the temporal stability of personality by questionnaire. *Journal of Personality and Social Psychology* **56**, 777–783.
- SHIFMAN, S., BHOMRA, A., SMILEY, S., WRAY, N. R., JAMES, M. R., MARTIN, N. G., HETTEMA, J. M., AN, S. S., NEALE, M. C., VAN DEN OORD, E., KENDLER, K. S., CHEN, X., BOOMSMA, D. I., MIDDELDORP, C. M., HOTTENGA, J. J., SLAGBOOM, P. E. & FLINT, J. (2008). A whole genome association study of neuroticism using DNA pooling. *Molecular Psychiatry* **13**, 302–312.
- SIH, A. & BELL, A. M. (2008). Insights for behavioral ecology from behavioral syndromes. In *Advances in the Study of Behavior, Vol 38, vol. 38. Advances in the Study of Behavior*, pp. 227–281.
- SIH, A., BELL, A. M. & JOHNSON, J. C. (2004a). Behavioral syndromes: an ecological and evolutionary overview. *Trends in Ecology and Evolution* **19**, 372–378.
- SIH, A., BELL, A. M., JOHNSON, J. C. & ZIEMBA, R. E. (2004b). Behavioral syndromes: An integrative overview. *Quarterly Review of Biology* **79**, 241–277.
- SIH, A., KATS, L. B. & MAURER, E. F. (2003). Behavioural correlations across situations and the evolution of antipredator behaviour in a sunfish-salamander system. *Animal Behaviour* **65**, 29–44.
- SINN, D. L., APIOLAZA, L. A. & MOLTSCHANIWSKYJ, N. A. (2006). Heritability and fitness-related consequences of squid personality traits. *Journal of Evolutionary Biology* **19**, 1437–1447.
- SINN, D. L., GOSLING, S. D. & MOLTSCHANIWSKYJ, N. A. (2008). Development of shy/bold behaviour in squid: context-specific phenotypes associated with developmental plasticity. *Animal Behaviour* **75**, 433–442.
- 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**, 99–110.
- SLUYTER, F., VANDERVLUGT, J. J., VANOORTMERSEN, G. A., KOOLHAAS, J. M., VANDERHOEVEN, F. & DEBOER, P. (1996). Studies on wild house mice. VII. Prenatal maternal environment and aggression. *Behavior Genetics* **26**, 513–518.
- SMISETH, P. T., WRIGHT, J. & KOLLIKER, M. (2008). Parent-offspring conflict and co-adaptation: behavioural ecology meets quantitative genetics. *Proceedings of the Royal Society B-Biological Sciences* **275**, 1823–1830.
- SMITH, B. R. & BLUMSTEIN, D. T. (2008). Fitness consequences of personality: a meta-analysis. *Behavioral Ecology* **19**, 448–455.
- STAMPS, J. (2003). Behavioural processes affecting development: Tinbergen's fourth question comes of age. *Animal Behaviour* **66**, 1–13.
- STAMPS, J. A. (2007). Growth-mortality tradeoffs and 'personality traits' in animals. *Ecology Letters* **10**, 355–363.

- SZYF, M., MCGOWAN, P. & MEANEY, M. J. (2008). The social environment and the epigenome. *Environmental and Molecular Mutagenesis* **49**, 46–60.
- TAYLOR, P. D. & DAY, T. (2004). Stability in negotiation games and the emergence of cooperation. *Proceedings of the Royal Society of London Series B-Biological Sciences* **271**, 669–674.
- TU, Y. K. & GILTHORPE, M. S. (2007). Revisiting the relation between change and initial value: A review and evaluation. *Statistics in Medicine* **26**, 443–457.
- UHER, J., ASENDORPF, J. B. & CALL, J. (2008). Personality in the behaviour of great apes: temporal stability, cross-situational consistency and coherence in response. *Animal Behaviour* **75**, 99–112.
- ULLER, T. (2008). Developmental plasticity and the evolution of parental effects. *Trends in Ecology & Evolution* **23**, 432–438.
- VAN HIERDEN, Y. M., KORTE, S. M., RUESINK, E. W., VAN REENEN, C. G., ENGEL, B., KORTE-BOUWS, G. A. H., KOOLHAAS, J. M. & BLOKHUIS, H. J. (2002). Adrenocortical reactivity and central serotonin and dopamine turnover in young chicks from a high and low feather-pecking line of laying hens. *Physiology & Behavior* **75**, 653–659.
- VAN OERS, K., DE JONG, G., DRENT, P. J. & VAN NOORDWIJK, A. J. (2004). A genetic analysis of avian personality traits: Correlated, response to artificial selection. *Behavior Genetics* **34**, 611–619.
- VAN OERS, K., DE JONG, G., VAN NOORDWIJK, A. J., KEMPENAEERS, B. & DRENT, P. J. (2005a). Contribution of genetics to the study of animal personalities: a review of case studies, pp. 1185–1206.
- VAN OERS, K., KLUNDER, M. & DRENT, P. J. (2005b). Context dependence of personalities: risk-taking behavior in a social and a nonsocial situation. *Behavioral Ecology* **16**, 716–723.
- VEENEMA, A. H., KOOLHAAS, J. M. & DE KLOET, E. R. (2004). Basal and stress-induced differences in HPA axis, 5-HT responsiveness, and hippocampal cell proliferation in two mouse lines. In *Stress: Current Neuroendocrine and Genetic Approaches*, vol. 1018. *Annals of the New York Academy of Sciences* (ed. K. Pacak, B. Aguilera, E. Saban and R. Kvetnansky), pp. 255–265.
- VERBEEK, M. E. M., DRENT, P. J. & WIEPKEMA, P. R. (1994). Consistent individual differences in early exploratory behavior of male Great Tits. *Animal Behaviour* **48**, 1113–1121.
- WEBSTER, M. M., WARD, A. J. W. & HART, P. J. B. (2007). Boldness is influenced by social context in threespine sticklebacks (*Gasterosteus aculeatus*). *Behaviour* **144**, 351–371.
- WEISS, A., KING, J. E. & HOPKINS, W. D. (2007). A cross-setting study of chimpanzee (*Pan troglodytes*) personality structure and development: Zoological parks and Yerkes National Primate Research Center. *American Journal of Primatology* **69**, 1264–1277.
- WEST-EBERHARD, M. J. (2003). *Developmental Plasticity and Evolution*. Oxford University Press, New York.
- WILLIAMS, T. D. (2008). Individual variation in endocrine systems: moving beyond the ‘tyranny of the Golden Mean’. *Philosophical Transactions of the Royal Society B-Biological Sciences* **363**, 1687–1698.
- WILSON, D. S., CLARK, A. B., COLEMAN, K. & DEARSTYNE, T. (1994). Shyness and boldness in humans and other animals. *Trends in Ecology & Evolution* **9**, 442–446.
- WILSON, D. S., COLEMAN, K., CLARK, A. B. & BIEDERMAN, L. (1993). Shy bold continuum in Pumpkinseed Sunfish (*Lepomis gibbosus*) – an ecological study of a psychological trait. *Journal of Comparative Psychology* **107**, 250–260.
- WOLF, M., VAN DOORN, G. S., LEIMAR, O. & WEISSING, F. J. (2007). Life-history trade-offs favour the evolution of animal personalities. *Nature* **447**, 581–584.
- WOLF, M., VAN DOORN, G. S. & WEISSING, F. J. (2008). Evolutionary emergence of responsive and unresponsive personalities. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 15825–15830.
- ZHOU, Y. H., KUSTER, H. K., PETTIS, J. S., DANKA, R. G., GLEASON, J. M. & GREENFIELD, M. D. (2008). Reaction norm variants for male calling song in populations of *Achroia grisella* (Lepidoptera: Pyralidae): Toward a resolution of the lek paradox. *Evolution* **62**, 1317–1334.